

STIC-Biotech/ChemLib

112 464

**From:** Whiteman, Brian  
**Sent:** Wednesday, January 21, 2004 8:05 AM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Fw: seq search

RECEIVED  
JAN 21 2004  
STIC

Labelled mailbox and room number

09/807,802  
Wilson et al., 11/29/01

search SEQ ID NOs: 13, 15, and 17 against us patent and us patent application databases.

the sequences are AAV-1 capsid proteins

Thanks,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

-----Original Message-----

**From:** Whiteman, Brian  
**Sent:** Wednesday, January 21, 2004 7:36 AM  
**To:** STIC-Biotech/ChemLib  
**Subject:** updated seq search

09/807,802  
Wilson et al., 11/29/01

search SEQ ID NOs: 13, 15, and 17 against us patent and us patent application databases.

the sequences are AAV-1 capsid proteins

Thank you,  
Brian Whiteman, 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
Remsen, 2D14  
(703) 305-0775

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:00:06 ; Search time 38.198 Seconds  
(without alignments)  
4972.168 Million cell updates/sec

Title: US-09-807-802a-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWLENDLSEGR.....NNGLYTEPRPIGTLYLTREL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phase:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriaph:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3989	100.0	736	12	Q9WBP8 adeno-associ
2	3963	99.3	736	12	O56137 adeno-associ
3	3511	88.0	736	12	O56139 adeno-associ
4	3494	87.6	736	12	O56311 adeno-associ
5	3442.5	86.3	737	12	O8UQGO adeno-associ
6	3405	85.4	738	12	O8UQF8 adeno-associ
7	3402.5	85.3	735	12	O56652 adeno-associ
8	2759.5	69.2	598	12	O56653 adeno-associ
9	2486.5	62.3	734	12	O41855 adeno-associ
10	2481.5	62.2	533	12	O92917 adeno-associ
11	2220	55.7	724	12	O9YI11 adeno-associ
12	2190.5	54.9	732	12	O83290 muscovy duc
13	2173.5	54.5	732	12	Q67666 goose parvo
14	2168.5	54.4	732	12	O83289 muscovy duc
15	2165.5	54.3	732	12	O8V395 goose parvo
16	2162.5	54.2	732	12	Q65444 barbarie du

17	1990.5	49.9	676	12	Q67672	Q67672 goose parvo
18	1712	42.9	587	12	O67667	Q67667 goose parvo
19	1698	42.6	534	12	O67668	Q67668 goose parvo
20	1695	42.5	587	12	Q65445	Q65445 barbarie du
21	1676	42.0	534	12	Q65446	Q65446 barbarie du
22	859.5	21.5	703	12	Q8QQV5	Q8QQV5 minute viru
23	715	17.9	179	12	Q9W8U7	Q9W8U7 duck parvo
24	715	17.9	179	12	Q9WA24	Q9WA24 goose parvo
25	711	17.8	179	12	Q9WN18	Q9WN18 duck parvo
26	709	17.8	179	12	Q9WN19	Q9WN19 duck parvo
27	707	17.7	179	12	Q9WN20	Q9WN20 goose parvo
28	618	15.5	571	12	Q8QQV4	Q8QQV4 minute viru
29	617	15.5	785	12	Q9J0X4	Q9J0X4 pig-tailed
30	614.5	15.4	781	12	Q9PZT0	Q9PZT0 human parvo
31	614	15.4	781	12	Q8JN56	Q8JN56 human parvo
32	611	15.3	981	12	Q8QVLI	Q8QVLI hamster par
33	608.5	15.3	781	12	Q9JGP8	Q9JGP8 human parvo
34	606.5	15.2	781	12	Q89318	Q89318 human parvo
35	606.5	15.2	781	12	Q912B8	Q912B8 human eryth
36	606.5	15.2	781	12	Q89319	Q89319 human parvo
37	606.5	15.2	781	12	Q89317	Q89317 human parvo
38	605.5	15.2	781	12	Q8JYE3	Q8JYE3 erythroviru
39	605	15.2	781	12	Q8JYD9	Q8JYD9 erythroviru
40	604.5	15.2	773	12	Q913X1	Q913X1 human parvo
41	604.5	15.2	781	12	Q89316	Q89316 human parvo
42	603.5	15.1	781	12	P90223	P90223 human parvo
43	603.5	15.1	781	12	P90221	P90221 human parvo
44	603.5	15.1	781	12	P90224	P90224 human parvo
45	603.5	15.1	781	12	Q85117	Q85117 human parvo

ALIGNMENTS

RESULT 1

Q9WBP8	PRELIMINARY;	PRT;	736 AA.
AC Q9WBP8;			
DT 01-NOV-1999 (Tremblrel. 12, Created)			
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)			
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)			
DE Capsid protein.			
OS Adeno-associated virus 1.			
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.			
OX NCBI_TaxID=85106;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=99214338; PubMed=10196295;			
RA Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;			
RT "Gene therapy vectors based on adeno-associated virus type 1.";			
RL J. Virol. 73:3994-4003(1999).			
RN [2]			
RP SEQUENCE FROM N.A.			
RA Xiao W., Wilson J.M.;			
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AF063497; AAD2757.1; -			
DR InterPro; IPR001403; Parvo.coat.			
DR Pfam; PF00740; Parvo.coat.1			
SQ SEQUENCE 736 AA; 81375 MW; CFABF9BD5CD0595 CRC64;			
Query Match	100.0%;	Score 3989;	DB 12; Length 736;
Best Local Similarity	100.0%;	Pred. No. 2e-266;	
Matches 736;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY 1	MAADGYLPDWLENDLSEGRWDLKPGAPKPKANQKQDDGRGLVLPQYKYLGPFGNGLD 60		
Db			
QY 1	MAADGYLPDWLENDLSEGRWDLKPGAPKPKANQKQDDGRGLVLPQYKYLGPFGNGLD 60		
Db			
QY 61	KGEFVNAADAALHDKAYDQOLKAGDNPYLRYNHADAETQERLQEDTSFGNGLGRAVFQ 120		
Db			
QY 61	KGEFVNAADAALHDKAYDQOLKAGDNPYLRYNHADAETQERLQEDTSFGNGLGRAVFQ 120		
Db			
QY 121	AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGTKGQQQPAKRLNFGQTGDSE 180		

[illegible]

```

RN  [2]
RP  SEQUENCE FROM N.A.
RA  Rutledge E.A., Russell D.W.;
RL  Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF028705; AAB95452.1; -
DR  InterPro; IPR001403; Parvo_coat.
DR  Pfam; PF00740; Parvo_coat; 1.
SQ  SEQUENCE 736 AA; 81906 MW; DD523331AD5FD070F CRC64;

Query Match      88.0%; Score 3511; DB 12; Length 736;
Best Local Similarity 86.8%; Pred. No. 1.8e-233;
Matches 640; Conservative 38; Mismatches 57; Indels 2; Gaps 2;

QY  1 MAADGYLPDWLENDLSEGIKREWWDLKPGAPKPKANQKQDDGRGLVLPYKYLGPNGLD 60
DB  1 MAADGYLPDWLENDLSEGIKREWWDLKPGAPKPKANQKQDDGRGLVLPYKYLGPNGLD 60
QY  61 KGEPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAVFQ 120
DB  61 KGEPVNEADAAALEHDKAYDQOLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAVFQ 120
QY  121 AKKRVLLEPLGLVEEGAKTAPGKCRPVQEQSPQSDSSGIGKGTQQAQKRLNFGQDSE 180
DB  121 AKKRILEPLGLVEEAAKTAPGKCRPVQEQSPQSDSSGIGKGTQQAQKRLNFGQDSE 180
QY  181 SVDPDQPLGEPPATPAAGVPTTWSGGGAPMADNNEGADGVGNSSGNWHCDSQWLGRDVI 240
DB  181 SVDPDQPLGEPPATPAAGVPTTWSGGGAPMADNNEGADGVGNSSGNWHCDSQWLGRDVI 240
QY  241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPNGYDFNRFCHFSPRDQRL 300
DB  241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPNGYDFNRFCHFSPRDQRL 300
QY  301 INNWGFRPKRLNFKLNIQVKEVTNDGVTIANNLTSTVQVFSDEYQLPVLSAQH 360
DB  301 INNWGFRPKLSFKLNIQVKEVTNDGVTIANNLTSTVQVFSDEYQLPVLSAQH 360
QY  361 GCLPPFPADVFMIPQGYLTLNNGSQAVGRSSFCYCLEYFPPSQMLRTGNNFTFSYTFEYVP 420
DB  361 GCLPPFPADVFMIPQGYLTLNNGSQAVGRSSFCYCLEYFPPSQMLRTGNNFTFSYTFEYVP 420
QY  421 FHSSYAHQSGLDRMLNPLIDQYLYLNRTQ-NQSGAQNKDLSFRSGSPAGMSVQPKNL 479
DB  421 FHSSYAHQSGLDRMLNPLIDQYLYLNRTQ-NQSGAQNKDLSFRSGSPAGMSVQPKNL 479
QY  480 PGFCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKPPMSG 539
DB  480 PGFCYRQRLSKTANDNNNSNFTWTGASKYNLNGRDSLVNPGPAMASHKDDKPPMSG 539
QY  540 VMI FGKESAGASNTALDNVMTDEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAM 599
DB  540 NLI FGKEGTTASNAELDNVMTDEBEIRTNPNVATEQGYTVANNLQSSNTAPTTRTVNDQ 599
QY  600 GALPGWVQDRDVLQGPWIAKIPHTDGHFHPSPLMGGGLKNPPQIILKNTVPANPP 659
DB  600 GALPGWVQDRDVLQGPWIAKIPHTDGHFHPSPLMGGGLKHPPIQIMIKNTVPANPP 659
QY  660 AEFSAFKASFTQYSTGQVSVEIWELOKENSKEWNPVQYTSNTAKSANVDFTVDNNG 719
DB  660 TTFSPAKFASFTQYSTGQVSVEIWELOKENSKEWNPQIYTSNTNKSINVVDFTVDNNG 719

```

## RESULT 4

```

Q65311 PRELIMINARY; PRT; 736 AA.
ID AC
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)

```

```

DE  Capsid protein.
OS  Adeno-associated virus 3.
OC  Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX  NCBI_TaxID=46350;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Muramatsu S., Brown K.E.;
RX  MEDLINE=96266430; PubMed=8661429;
RL  Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR  EMBL; U48704; AAC55049.1; -
DR  InterPro; IPR001403; Parvo_coat.
DR  Pfam; PF00740; Parvo_coat; 1.
SQ  SEQUENCE 736 AA; 81660 MW; AFF1EF47B5C67A10 CRC64;

Query Match      87.6%; Score 3494; DB 12; Length 736;
Best Local Similarity 86.4%; Pred. No. 2.6e-232;
Matches 637; Conservative 39; Mismatches 59; Indels 2; Gaps 2;

QY  1 MAADGYLPDWLENDLSEGIKREWWDLKPGAPKPKANQKQDDGRGLVLPYKYLGPNGLD 60
DB  1 MAADGYLPDWLENDLSEGIKREWWDLKPGAPKPKANQKQDDGRGLVLPYKYLGPNGLD 60
QY  61 KGEPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAVFQ 120
DB  61 KGEPVNEADAAALEHDKAYDQOLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAVFQ 120
QY  121 AKKRVLLEPLGLVEEGAKTAPGKCRPVQEQSPQSDSSGIGKGTQQAQKRLNFGQDSE 180
DB  121 AKKRILEPLGLVEEAAKTAPGKCRPVQEQSPQSDSSGIGKGTQQAQKRLNFGQDSE 180
QY  181 SVDPDQPLGEPPATPAAGVPTTWSGGGAPMADNNEGADGVGNSSGNWHCDSQWLGRDVI 240
DB  181 SVDPDQPLGEPPATPAAGVPTTWSGGGAPMADNNEGADGVGNSSGNWHCDSQWLGRDVI 240
QY  241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPNGYDFNRFCHFSPRDQRL 300
DB  241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPNGYDFNRFCHFSPRDQRL 300
QY  301 INNWGFRPKRLNFKLNIQVKEVTNDGVTIANNLTSTVQVFSDEYQLPVLSAQH 360
DB  301 INNWGFRPKLSFKLNIQVKEVTNDGVTIANNLTSTVQVFSDEYQLPVLSAQH 360
QY  361 GCLPPFPADVFMIPQGYLTLNNGSQAVGRSSFCYCLEYFPPSQMLRTGNNFTFSYTFEYVP 420
DB  361 GCLPPFPADVFMIPQGYLTLNNGSQAVGRSSFCYCLEYFPPSQMLRTGNNFTFSYTFEYVP 420
QY  421 FHSSYAHQSGLDRMLNPLIDQYLYLNRTQ-NQSGAQNKDLSFRSGSPAGMSVQPKNL 479
DB  421 FHSSYAHQSGLDRMLNPLIDQYLYLNRTQ-NQSGAQNKDLSFRSGSPAGMSVQPKNL 479
QY  480 PGFCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKPPMSG 539
DB  480 PGFCYRQRLSKTANDNNNSNFTWTGASKYNLNGRDSLVNPGPAMASHKDDKPPMSG 539
QY  540 VMI FGKESAGASNTALDNVMTDEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAM 599
DB  540 NLI FGKEGTTASNAELDNVMTDEBEIRTNPNVATEQGYTVANNLQSSNTAPTTRTVNDQ 599
QY  600 GALPGWVQDRDVLQGPWIAKIPHTDGHFHPSPLMGGGLKNPPQIILKNTVPANPP 659
DB  600 GALPGWVQDRDVLQGPWIAKIPHTDGHFHPSPLMGGGLKHPPIQIMIKNTVPANPP 659
QY  660 AEFSAFKASFTQYSTGQVSVEIWELOKENSKEWNPVQYTSNTAKSANVDFTVDNNG 719
DB  660 TTFSPAKFASFTQYSTGQVSVEIWELOKENSKEWNPQIYTSNTNKSINVVDFTVDNNG 719

```



```
QY 720 LYTEPRPIGTRYLTRPL 736
Db 720 VYSEPRPIGTRYLTRNL 736

RESULT 5
Q8JQGO PRELIMINARY; PRT; 737 AA.
AC Q8JQGO;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DE Capaid protein.
OS Adeno-associated virus 7.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=202812;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12192090;
RA Gao G.P., Alvira M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
human gene therapy."
RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859 (2002).
DR EMBL; AF513851; AAO03855.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 737 AA; 81652 MW; D63B8CE7583140D0 CRC64;

Query Match 86.3%; Score 3442.5; DB 12; Length 737;
Best Local Similarity 85.3%; Pred. No. 9.4e-229;
Matches 630; Conservative 50; Mismatches 54; Indels 5; Gaps 4;

QY 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQKQDDGRGLVLPYKYLGFPGNGLD 60
Db 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQKQDDGRGLVLPYKYLGFPGNGLD 60

QY 61 KGEVNAADAAALEHDKAYDQQLKAGDNPYLYRNHADADEFQERLQEDTSFGCNLGRAVQ 120
Db 61 KGEVNAADAAALEHDKAYDQQLKAGDNPYLYRNHADADEFQERLQEDTSFGCNLGRAVQ 120

QY 121 AKRVLEPLGLVEEGAKTAPGKRVPESQPO-EPDSSSGIGKTGQOPAKKRLNFGQDGS 179
Db 121 AKRVLEPLGLVEEGAKTAPGKRVPESQPO-EPDSSSGIGKTGQOPAKKRLNFGQDGS 180

QY 180 ESVDPDQPLGEPPTAAVGPPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRV 239
Db 181 ESVDPDQPLGEPPTAAVGPPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRV 240

QY 240 ITTSTRTWALPTNNHLYKQISSASTGSANDNHYFGYSTPMGYDFNRFCHFSPRDQW 299
Db 241 ITTSTRTWALPTNNHLYKQISSASTGSANDNHYFGYSTPMGYDFNRFCHFSPRDQW 300

QY 300 LINNNGFRPKRLNFKLFIQVKEVTNDGVTTIANNLTSTQVFSDESQYLPVYLGSAH 359
Db 301 LINNNGFRPKRLNFKLFIQVKEVTNDGVTTIANNLTSTQVFSDESQYLPVYLGSAH 360

QY 360 QGCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPSSOMLRTGNFTFSYTFEEV 419
Db 361 QGCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPSSOMLRTGNFTFSYTFEEV 420

QY 420 PFHSSYAHQSQSLDRMLNPLIDQYLYLNRTQ-NQSGAONKDLLPSRGSAGMSVQPKRW 478
Db 421 PFHSSYAHQSQSLDRMLNPLIDQYLYLNRTQ-NQSGAONKDLLPSRGSAGMSVQPKRW 480

QY 479 LQPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMS 538
Db 481 LQPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMS 540

QY 539 GVMIFKESAGASN-TALDNVMTIDBEEIKATNPVATERFGTVAVNFQSSSDPATGDVH 597
Db 541 GVLIFCK--TGATNKTTLLENVMTNEEELRPINPVATEEYGVSSNLQAANTAAQTQVNV 598
```

```
QY 598 AMGALPGVMQDRDYLQGPITWAKIPHTDGHFHPSPLMGFGFLGNPPQILIKNTVPAN 657
Db 599 NQOGLPGVMQDRDYLQGPITWAKIPHTDGHFHPSPLMGFGFLGNPPQILIKNTVPAN 658

QY 658 PPAEFSATKPAFPTOYSTQGVSVIEIWELOKENSKRWNPEVOYTSNYAKSANVDFTVDN 717
Db 659 PPEVFTPAKFASFITQYSTQGVSVIEIWELOKENSKRWNPEIQTNSPEKOTGVDFAVDS 718

QY 718 NGLYTEPRPIGTRYLTRPL 736
Db 719 QGVSEPRPIGTRYLTRNL 737

RESULT 6
Q8JQF8 PRELIMINARY; PRT; 738 AA.
AC Q8JQF8;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Capaid protein.
OS Adeno-associated virus 8.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=202813;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12192090;
RA Gao G.P., Alvira M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
human gene therapy."
RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859 (2002).
DR EMBL; AF513852; AAO03857.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 738 AA; 81756 MW; 59BC2BE1B148AA14 CRC64;

Query Match 85.4%; Score 3405; DB 12; Length 738;
Best Local Similarity 84.0%; Pred. No. 3.6e-226;
Matches 620; Conservative 48; Mismatches 68; Indels 2; Gaps 2;

QY 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQKQDDGRGLVLPYKYLGFPGNGLD 60
Db 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQKQDDGRGLVLPYKYLGFPGNGLD 60

QY 61 KGEVNAADAAALEHDKAYDQQLKAGDNPYLYRNHADADEFQERLQEDTSFGCNLGRAVQ 120
Db 61 KGEVNAADAAALEHDKAYDQQLKAGDNPYLYRNHADADEFQERLQEDTSFGCNLGRAVQ 120

QY 121 AKRVLEPLGLVEEGAKTAPGKRVPESQPO-EPDSSSGIGKTGQOPAKKRLNFGQDGS 179
Db 121 AKRVLEPLGLVEEGAKTAPGKRVPESQPO-EPDSSSGIGKTGQOPAKKRLNFGQDGS 180

QY 180 ESVDPDQPLGEPPTAAVGPPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRV 239
Db 181 ESVDPDQPLGEPPTAAVGPPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRV 240

QY 240 ITTSTRTWALPTNNHLYKQISS-ASTGASNDNHYFGYSTPMGYDFNRFCHFSPRDQW 298
Db 241 ITTSTRTWALPTNNHLYKQISS-ASTGASNDNHYFGYSTPMGYDFNRFCHFSPRDQW 300

QY 299 RLINNNGFRPKRLNFKLFIQVKEVTNDGVTTIANNLTSTQVFSDESQYLPVYLGSA 358
Db 301 RLINNNGFRPKRLNFKLFIQVKEVTNDGVTTIANNLTSTQVFSDESQYLPVYLGSA 360

QY 359 HQGCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPSSOMLRTGNFTFSYTFEE 418
Db 361 HQGCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPSSOMLRTGNFTFSYTFEE 420

QY 419 VPFHSSYAHQSQSLDRMLNPLIDQYLYLNRTQ-NQSGAONKDLLPSRGSAGMSVQPKRW 478
Db 421 VPFHSSYAHQSQSLDRMLNPLIDQYLYLNRTQ-NQSGAONKDLLPSRGSAGMSVQPKRW 480

QY 479 LQPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMS 538
```

```

Db 481 LPGPCYRQVRSTTTGQNNNSFEATAGTKYHLNGRSLANPGIATAMATHKDDBERFFPSN 540
QY 539 GVMIKESAGANTALDNVMTDEBEIKATNPVATERFCTVAVNFQSSSTDPATGDVHA 598
Db 541 GILIFGKQNAARDNADSYDNLVTSEBEIKTNPVATEEYGIADVADNLQQQNTAPQIGTVNS 600
QY 599 MGALPGVMQDRDYLQGPWAKIPIHTDGHFHPSPMLMGFGGLKPNPPPPQILIKNTVPANP 658
Db 601 QGALPGVMQNRDYLQGPWAKIPIHTDGNFHPSPMLMGFGGLKHPPPQILIKNTVPADP 660
QY 659 PAFSATKPAFPITQYSTGVSVIEIWELOKENSKRWNPEVQTSNYAKSANYDFTVDNN 718
Db 661 PTTFNQSKLSFITYSTGVSVIEIWELOKENSKRWNPEIQYTSNYKTSYSDFAVNT 720
QY 719 GLYTEPRPIGTRYLRL 736
Db 721 GVTSEPRPIGTRYLRL 738

RESULT 7
O56652 PRELIMINARY; PRT; 735 AA.
AC O56652;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major coat protein VP1.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF043303; AAC03780.1; --
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
KW Coat protein.
FT VARIANT 76 76 D -> V.
FT VARIANT 553 553 D -> N.
FT VARIANT 567 567 T -> N.
FT VARIANT 677 678 QV -> HV.
FT VARIANT 710 710 V -> R.
SQ SEQUENCE 735 AA; 81944 MW; 980BEEF46908390B CRC64;

Query Match 85.3%; Score 3402.5; DB 12; Length 735;
Best Local Similarity 83.3%; Pred. No. 5.3e-226;
Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPDWLENDLSEGIREWDLKPGAPKPKANQKQDDGRGLVLPQYKYLGPNGLD 60
Db 1 MAADGYLPDWLEDTLSEGIQWMLKPGPPPPKPAERHKDDSRGLVLPQYKYLGPNGLD 60
QY 61 KGPVNAADAALHDKAYDQQLKAGDNPYLYRNHDAEFQERLQEDTSFGNGLGRAVFQ 120
Db 61 KGPVNEADAALHDKAYDRQLDSGDNPLYLYRNHDAEFQERLQEDTSFGNGLGRAVFQ 120
QY 121 AKKRVLEPLGLVEEGATPAKRPVPSQSPQSDSSGIQKTKQPAKRLNFGQTGDSE 180
Db 121 AKKRVLEPLGLVEEVPKTAFGKRPVSHSVPEDSSSGTGKQGPARKRLNFGQTGDAD 180
QY 181 SVDPDQPLGEPPTAPAAVGTPTWASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240

```

```

Db 181 SVDPDQPLGEPPTAPAAVGTPTWATGSGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240
QY 241 TTSTRTWALPTNNHLYKQISSASTGASNDNHFGYSTPWGYPDENRFCHESPRDWORL 300
Db 241 TTSTRTWALPTNNHLYKQISSQS-GASNDNHFGYSTPWGYPDENRFCHESPRDWORL 299
QY 301 INNNMGRFKRLNFKLFIQVKEVTNDGVTIANNLTSTVQVFSDESQYLPVVLGSAHQ 360
Db 300 INNNMGRFKRLNFKLFIQVKEVTNDGVTIANNLTSTVQVFTDSEYQLPVVLGSAHQ 359
QY 361 GCLPPPPADVPMIPQGYITLNGSOAVGRSSFYCLEYFPSSQMLRTGNNFTSYTPEEVP 420
Db 360 GCLPPPPADVPMIPQGYITLNGSOAVGRSSFYCLEYFPSSQMLRTGNNFTSYTPEEVP 419
QY 421 FHSSVAHSOSLDRLMNLIDQYLYLNRNTONSGAQNKDLLFSRGSAGMSVQPKNWLP 480
Db 420 FHSSVAHSOSLDRLMNLIDQYLYLSRTNTVSGTTTQSRLOFSQAGASDIRDSRWLP 479
QY 481 GPCYRQVRVSKTKTNNNSNFTWTGASKYLNLRGSIINPGTAMASHKDDDEKFFPMGV 540
Db 480 GPCYRQVRVSKTSADNNSEYSWTGATKYHLNGRDSLVPNPGPAMASHKDDDEKFFPMGV 539
QY 541 MIFGESAGANTALDNVMTDEBEIKATNPVATERFCTVAVNFQSSSTDPATGDVHAMG 600
Db 540 LIFGKGSEKTNVDIEKVMITDEBEIRITNPVATEQYGSVSTNLQRGNRQAATADYNTQG 599
QY 601 ALPGVMQDRDYLQGPWAKIPIHTDGHFHPSPMLMGFGGLKPNPPPPQILIKNTVPANPPA 660
Db 600 VLPGMQDRDYLQGPWAKIPIHTDGHFHPSPMLMGFGGLKHPPPQILIKNTVPANPST 659
QY 661 EFSATKFAFITYSTGVSVIEIWELOKENSKRWNPEVQTSNYAKSANYDFTVDNGL 720
Db 660 TFSAAKFAFITYSTGVSVIEIWELOKENSKRWNPEIQYTSNYKSNVVDFTVDNGV 719
QY 721 YTEPRPIGTRYLRL 736
Db 720 YSEPRPIGTRYLRL 735

RESULT 8
O56653 PRELIMINARY; PRT; 598 AA.
AC O56653;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major coat protein VP2.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF043303; AAC03778.1; --
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
KW Coat protein.
FT VARIANT 416 416 D -> N.
FT VARIANT 430 430 T -> N.
FT VARIANT 540 541 QV -> HV.
FT VARIANT 573 573 V -> R.

```

```

SQ SEQUENCE 598 AA; 66619 MW; 070811ED9368E934 CRC64;
Query Match
Best Local Similarity 82.8%; Score 2759.5; DB 12; Length 598;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;
QY 139 AFGKRPVEQSPQEPDSSGIGKTQQQPAKRLNFGQDSDSVPDPPLGEPPTAAV 198
Db 2 AFGKRPVEHSVPEPDSSGCTKAGQQAQKRLNFGQDSDSVPDPPLGQPPAASGL 61
QY 199 GPTWASGGAPKADNNEGADGVGNHWCSTWLGDRVLTSTRWALPTNNHLYK 258
Db 62 GNTWATGSGAPKADNNEGADGVGNHWCSTWMDRVLTSTRWALPTNNHLYK 121
QY 259 QISSASTGASNDNHFGYTPMGYFDNRFHCHFSRDMQRLNINNWGFRPKRLNFKFN 318
Db 122 QISSQS-GASNDNHFGYSTPMGYFDNRFHCHFSRDMQRLNINNWGFRPKRLNFKFN 180
QY 319 IQKVEVTTNDGVTTIANNLTSTVQVPSSEYQLPYVLGSAHQCLPPPPADVFMIPOGY 378
Db 181 IQKVEVTTNDGVTTIANNLTSTVQVPSSEYQLPYVLGSAHQCLPPPPADVFMIPOGY 240
QY 379 LTLNNGSAVGRSSFYCLEYFSPQMLRTGNNFTESYTFEEVPHSGSYAHSQSLDRLMNP 438
Db 241 LTLNNGSAVGRSSFYCLEYFSPQMLRTGNNFTESYTFEEVPHSGSYAHSQSLDRLMNP 300
QY 439 IDOYLNNRTQNGSQAQNKDILLFSGSPAGMSVQPKNWLPCPCYRQVRVSKTKTDNNN 498
Db 301 IDOYLNNRTQNGSQAQNKDILLFSGSPAGMSVQPKNWLPCPCYRQVRVSKTKTDNNN 360
QY 499 SNPTWTGSKYNLNGRESINFGTAMASHKDEKFPFPMGSMVIFGKESAGASNTALDNV 558
Db 361 SEYSWTGATKYLNGRDSLNVNPGPAMASHKDEKFPFQSGVLIIFGKQSEKTNVDIEKV 420
QY 559 MITDEEIKATNPVATERGTAVNFOSSSTDPATGDVHAMGALPGMWQDRDVLQGP 618
Db 421 MITDEEIKATNPVATERGTAVNFOSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480
QY 619 WAKIPHTDGHFHPSPLMGGFGLKNNPPQILIKNTVPANPPAEFSATKFASTIQTSTQ 678
Db 481 WAKIPHTDGHFHPSPLMGGFGLKNNPPQILIKNTVPANPPAEFSATKFASTIQTSTQ 540
QY 679 VSVEIEWELQKSKRNWPEVQTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
Db 541 VSVEIEWELQKSKRNWPEVQTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 598

RESULT 9
O41855 PRELIMINARY; PRT; 734 AA.
AC O41855;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE Capsid.
OS Adeno-associated virus 4.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=57579;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC VR-646;
RX MEDLINE=97404695; PubMed=9261407;
RA Chiorini J.A., Yang L., Liu Y., Safer B., Kotin R.M.;
RT "Cloning of adeno-associated virus type 4 (AAV4) and generation of
recombinant AAV4 particles.";
RL J. Virol. 71:6823-6833(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC VR-646;
RA Chiorini J.A., Yang L., Kotin R.M., Safer B.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U89790; AAC58045.1; --
DR InterPro: IPR001403; Parvo_coat.

```

```

DR Pfam: PF00740; Parvo_coat; 1.
SQ SEQUENCE 734 AA; 80639 MW; 616CC27A777BBE6F CRC64;
Query Match
Best Local Similarity 62.3%; Score 2486.5; DB 12; Length 734;
Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;
QY 4 DGYLPDLWLEDNLSGIREWMDLKPAPKPKANQKQDQDGLVLPYGYKYLFPFNGLDKGE 63
Db 3 DGYLPDLWLEDNLSGIREWMDLKPAPKPKANQKQDQDGLVLPYGYKYLFPFNGLDKGE 62
QY 64 PYNAADAAALEHDKAYDQOLKAGDNPLYRYNHADAEPQERLOEDTSFGNIGRAVFAQKK 123
Db 63 PYNAADAAALEHDKAYDQOLKAGDNPLYRYNHADAEPQERLOEDTSFGNIGRAVFAQKK 122
QY 124 RVLBPGLVVEGAKTAPGKRPVQSPQEPDSSGIGKTQQQPAKRLNFGQDSDSVPDPPLG 182
Db 123 RVLBPGLVVEGAKTAPGKRPVQSPQEPDSSGIGKTQQQPAKRLNFGQDSDSVPDPPLG 182
QY 183 PDQPLGEPPTAAVGTTPWASGGAPKADNNEGADGVGNHWCSTWLGDRVLTSTRWALPT 242
Db 183 PEGSTSG-----AMSDDESENRAAGAAVEGGQAGGVGNASGDWHDSTWSEGHVTTT 236
QY 243 STRTWALPTNNHLYKQISSASTGASNDNHFGYTPMGYFDNRFHCHFSRDMQRLNIN 302
Db 237 STRTWALPTNNHLYKRLGE-----SLOSNTYNGFSTPMGYFDNRFHCHFSRDMQRLN 292
QY 303 NNWFRPKRLNFKLFIQKVEVTTNDGVTTIANNLTSTVQVPSSEYQLPYVLGSAHQCL 362
Db 293 NNWGRPKAMRVKLFNIQKVEVTTNSGETTVANNLTSTVQVIFADSSYELPYVMDAGQSG 352
QY 363 LPPPPADVFMIPOGY-----LTLNNGSAVGRSSFYCLEYFSPQMLRTGNNFTESYTF 419
Db 353 LPPPPADVFMIPOGY-----LTLNNGSAVGRSSFYCLEYFSPQMLRTGNNFTESYTF 412
QY 420 PPHSYAHSQSLDRLMNPIDYOYLNNRTQNGSQAQNKDILLFSGSPAGMSVQPKNWL 475
Db 413 PPHSYAHSQSLDRLMNPIDYOYLNNRTQNGSQAQNKDILLFSGSPAGMSVQPKNWL 469
QY 476 KMWLPGPCYRQVRVSKTKTDNNNFTMGAS---KY----NLNGRESIINPGTAMASHK 528
Db 470 KMWLPGPCYRQVRVSKTKTDNNNFTMGAS---KY----NLNGRESIINPGTAMASHK 527
QY 529 DDEKFFPMGSMVIFGKESAGASNTALDNVITDDEEIKATNPVATERGTAVNFOSSST 588
Db 528 PADSK-FSNSQLIFAGPKQNGNTATVPGTLIFTSEELAATNATDTMMGNLPGGDQNS 586
QY 589 TDPATGDVHAMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPLMGGFGLKNNPPQIL 648
Db 587 NLPTVDRLTALGAVPGMWQDRDIYQGPWAKIPHTDGHFHPSPLMGGFGLKNNPPQIL 646
QY 649 IKNTVPANPPAEFSATKFASTIQTSTQVSVIEWELQKSKRNWPEVQTSNYAKS 708
Db 647 IKNTVPANPPAEFSATKFASTIQTSTQVSVIEWELQKSKRNWPEVQTSNYAKS 706
QY 709 ANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
Db 707 NSLLWAPDAAGKYTEPRPIGTRYLTRPL 734

RESULT 10
O92917 PRELIMINARY; PRT; 533 AA.
AC O92917;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE Major coat protein VP3.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=9508582; PubMed=7996133;  
RA Ruffing M., Heid H., Kleinschmidt J.A.;  
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid  
RT proteins affect viral infectivity: lack of an RGD integrin-binding  
RT motif".  
RL J. Gen. Virol. 75:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,  
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,  
RA Tratschin J.-D., Weitz M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043303; AAC03779.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein.  
FT VARIANT 351 351 D -> N.  
FT VARIANT 365 365 T -> N.  
FT VARIANT 475 476 QV -> HV.  
FT VARIANT 508 508 V -> R.  
SQ SEQUENCE 533 AA; 60063 MW; 9E4DB8C25810D4F0 CRC64;

Query Match 62.2%; Score 2481.5; DB 12; Length 533;  
Best Local Similarity 83.3%; Pred. No. 18-162;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
QY 203 MASGGAPMADNNEGADGVNAGSNHWCSTWLGDRVITTTSTRTWALPTNNHLYKQISS 262  
DB 1 MATSGAPMADNNEGADGVNAGSNHWCSTWLGDRVITTTSTRTWALPTNNHLYKQISS 60  
QY 263 ASTGASNDNHYGYSYTPWGYFDFNRPCHFSRDRWRLNNWGRFPRKLNFKLFNIQVK 322  
DB 61 QS-GASNDNHYGYSYTPWGYFDFNRPCHFSRDRWRLNNWGRFPRKLNFKLFNIQVK 119  
QY 323 EVTNDGVTTIANNTSTVQVESDSYQLPYVLGSAHQCLPFPADVFMIPQYGYLTIN 382  
DB 120 EVTNDGVTTIANNTSTVQVESDSYQLPYVLGSAHQCLPFPADVFMIPQYGYLTIN 179  
QY 383 NGSQAVGRSFFCYCLEYFPQMLRTGNNFTFSYFEEVFPFSSYAHSSQSLDRMLNPLIDQY 442  
DB 180 NGSQAVGRSFFCYCLEYFPQMLRTGNNFTFSYFEEVFPFSSYAHSSQSLDRMLNPLIDQY 239  
QY 443 LYLNTQNSGSAQNKDLFFSRGSPAGMSVQPKNLPGPCYRQRQSVKTKTDNNNSNFT 502  
DB 240 LYLSTRNTPSGTTQSRQLQFSQAGASDIRDQSRNLPGPCYRQRQSVKTSADNNNSYS 299  
QY 503 WTCGSKYNLNGRESIINPGTMASHKDDKDEKPPMSGVMI FGKESAGASNTALDNVMTD 562  
DB 300 WTCATKIHLNGRSLVNP GPAMASHKDDKDEKPPMSGVMI FGKQSGSEKTNVDIEKVMITD 359  
QY 563 EBEIKATNPVATRFGTAVNFGSSSTDPATGVDHAMGALPGMWQDRDVLQGPWAKI 622  
DB 360 EBEIRTNVATEQYGSVSTNLQRNGRQAADVNTQGVLPGMWQDRDVLQGPWAKI 419  
QY 623 PHTDGHFSPMLGGFGLKNPPQIILKNTVPANPPAFSATKPFASFTQYSTQGVSE 682  
DB 420 PHTDGHFSPMLGGFGLKNPPQIILKNTVPANPPAFSAAKPFASFTQYSTQGVSE 479  
QY 683 IEWELOKENSKRNPEVQVTSNYAKSANVDFTVDNGLYTEPRPIGTRYLTPL 736  
DB 480 IEWELOKENSKRNPEIQVTSNYKSNVNDFTVDNGLVSEPRPIGTRYLTPL 533

## RESULT 11

Q9YI1 PRELIMINARY; PRT; 724 AA.  
AC Q9YI1;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Capsid protein.  
GN CAP OR VP1.  
OS Adeno-associated virus 5.

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI TaxID=82300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=99098980; PubMed=9882294;  
RA Bantel-Schaal U., Deltus H., Schmidt R., Zur Hausen H.;  
RT "Human adeno-associated virus type 5 is only distantly related to  
RT other known primate helper-dependent parvoviruses".  
RL J. Virol. 73:939-947(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9909022; PubMed=9882336;  
RA Chlorini J.A., Kim F., Yang L., Kotin R.M.;  
RT "Cloning and characterization of adeno-associated virus type 5";  
RL J. Virol. 73:1309-1319(1999).  
DR EMBL; Y18065; CAA77024.1; -  
DR EMBL; AF085716; AAD13756.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 724 AA; 80424 MW; BBC32B88258D34C CRC64;

Query Match 55.7%; Score 2220; DB 12; Length 724;  
Best Local Similarity 58.6%; Pred. No. 1.7e-144;  
Matches 431; Conservative 81; Mismatches 198; Indels 26; Gaps 9;  
QY 8 PDWLEDNLSEGIRESWMDLPGAPKPKANQQQDDGRGLVLPFYKYLGPFGNGLDKGPEVNA 67  
DB 8 PDWLEE-VGEGLREFLGEAGPPKPKPNQHQDQARGLVLPVNYLGPNGLDRCPEVNR 66  
QY 68 ADAALAHDKAYDQOLKAGDNPYLYRNHADAEPQERLEDTSFGNGLGRAPVQAKRVLE 127  
DB 67 ADEVAREHDI SYNEQLEAGDNPYLYRNHADAEPQERLEDTSFGNGLKAVFQAKRVLE 126  
QY 128 PLGLVEEAKTAPGKRPVEQSPQESDSSSGIGTKGQAPKAKRLNFGOTGSESVDPDOP 187  
DB 127 PLGLVEEAKTAPGKRIIDHPPKPKKART-----EEDSKP-----STSSDAEAGPSSQ 176  
QY 188 LGEPPATPA-AVGPTTMASSGGAPMADNNEGADGVNAGSNHWCSTWLGDRVITTTSTRT 246  
DB 177 QLOIPAPASSLGADTMSAGGGPLGDNNGQADGVNAGSNHWCSTWMDRVVTKSTRT 236  
QY 247 WALPTYNNHLYKQISSASTGASNDNHYGYSTPWGYFDFNRPCHFSRDRWRLNNWNG 306  
DB 237 WYLPSEYNNHQYREIKSGSVDSGNANAYEGYSTPWGYFDFNRPCHFSRDRWRLNNWNG 296  
QY 307 FRPKELNFKLFNIQVKEVTTNDGVTTIANNTSTVQVESDSYQLPYVLGSAHQCLPFP 366  
DB 297 FRPKELNFKLFNIQVKEVTTNDGVTTIANNTSTVQVESDSYQLPYVLGSAHQCLPFP 356  
QY 367 PADVFMIPQYGYLTIN--NGSQA VGRSFFCYCLEYFPQMLRTGNNFTFSYFEEVFPFSS 424  
DB 357 PPQVPTLPQYGYATLNRDNTENPTERSFFCYCLEYFPQMLRTGNNFTFSYFEEVFPFSS 416  
QY 425 YAHSSQSLDRMLNPLIDQVLYLNTQNSGSAQNKDLFFSRGSPAGMSVQPKNLPGPC 483  
DB 417 FAPSQNLFKLANPLVDQYLYRFVSTNNTGGYQFNKVL-----AGRYANTYKMPFGPM 469  
QY 484 YRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTMASHKDDKDEKPPMSGVMI 543  
DB 470 GRTQWNVLGSGVNRASVSAPATTNRMELEGASYVPPQPNGTMTNQLQSNYTALENTMIF 529  
QY 544 GKESAGASNTAL---DNVMTDEBEIKATNPVATERFGTAVNFGSSSTDPATGVDHAMG 600  
DB 530 NSQAPANPGTTATYLEGNMLITSESETQPVNRVAVNVGGMATNQSSTTAPATGTYNLQE 589  
QY 601 ALPGMWQDRDVLQGPWAKI PHTDGHFSPMLGGFGLKNPPQIILKNTVPANPPAF 660  
DB 590 IVPGSMWMDRVDYLQGPWAKI PETGAHFHSPAGMGFGLKHPMPMMLIKNTVPVGN-IT 648  
QY 661 EFSATKPFASFTQYSTQGVSEIEWELOKENSKRNPEVQVTSNYAKSANVDFTVDNGL 720  
DB 649 SFSQVPSVSSFTQYSTQGVTVEMEWELKKNWNPFIQVTSNYKSNVNDPQVDFADPSTGE 708

```
QY 721 YTEPRPIGTRYLRL 736
Db 709 YRTTRPIGTRYLRL 724

RESULT 12
Q83290 PRELIMINARY; PRT; 732 AA.
AC Q83290;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Capsid protein.
OS Muscovy duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=37325;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=89384/France;
RX MEDLINE=96406928; PubMed=8811015;
RA Le Gall-Recule G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;
RT "Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)
in a baculovirus expression system and demonstration of immunity
induced by the recombinant protein.";
RL J. Gen. Virol. 77:2159-2163 (1996).
DR EMBL; Z68272; CAA32575.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
FT CHAIN 146 732 VP2 CAPSID PROTEIN.
FT CHAIN 199 732 VP3 CAPSID PROTEIN.
FT SEQUENCE 732 AA; 81364 MW; DE70DCCAB215F4E2 CRC64;

Query Match 54.9%; Score 2190.5; DB 12; Length 732;
Best Local Similarity 55.7%; Pred. No. 1.9e-142;
Matches 422; Conservative 88; Mismatches 184; Indels 63; Gaps 13;

QY 9 DWLEDNLSEGIREWMDLPGAPKPKANQKQ-----DDGRGLVLPQYKYLGP 57
Db 10 DWY-----TAAASWRHLKAGAPKPKSNQSQSVSTDRKPKDNNGRGFVLPQYKYLGP 65
QY 58 GLDKGPPVNAADAALHDKAYDQOLKAGDNPVLRYNHADAEFQERLQEDTSFGNGLGRA 117
Db 66 GLDKGPPVNAADSVALEHDKAYDQOLKAGDNPYIKFNHADQEFIDNLQDTSFGNGLGRA 125
QY 118 VFQAKRVLPLGLVEEGAKTAPGKRVPVEQSPQEPDSSSGIGTKTQGPAPKRLNFGQ 177
Db 126 VFQAKRILPLGLVEEPVNTAPAK-----SSG-KLTDHYPIVKKPKLSE-- 170
QY 178 DSESVPDPQPLG-----EPPATPAAGVPTTMASGGGAPMADNNEGADGVGNASG 226
Db 171 --ENSPPSNSSGGEASAAATEGSEPVAA-----NMAEGSGGAMGDSAGGADGVGNASG 222
QY 227 NHCDSTWLGDRVITSTRTWALPTNNHLYKQISSASTGASNDNHYFGYSTPWGYFDN 286
Db 223 NHCDSTWLGDRVITSTRTWALPTNNHLYKQISSASTGASNDNHYFGYSTPWGYFDN 281
QY 287 RFHCHFSRDMQRLNNHNGFRPKRLNFKLNIQVKEVTNDGVTIANNLSTVQVFS 346
Db 282 RFHCHFSRDMQRLNNHNGFRPKRLNFKLNIQVKEVTNDGVTIANNLSTVQVFS 341
QY 347 SEYQLPYVLGSAHQCLPFPADVFMIPQYGLTLN---NGSQAVRSFPCYLEYPPSQM 403
Db 342 NEHQLPYVLGSAHQCLPFPADVFMIPQYGLTLN---NGSQAVRSFPCYLEYPPSQM 401
QY 404 LRTGNFTSYTEPEEYFPHSSVAHSQSLDRMLNPLIDQYLYLNRNTQNSGAQNKDL 463
Db 402 LRTGNFTSYTEPEEYFPHSSVAHSQSLDRMLNPLIDQYLYLNRNTQNSGAQNKDL 455
QY 464 SRGSPAGMSVQPKNLPGPCYRQORVSKTK--TDNNNSNFTMTGASKYLNNGRESIN 521
Db 456 KKAVKGAFAFGMRNLPGPKLLDQVRAYSGGTDNVAANSKNGKFLKDRYLLQPG 515
QY 522 TAMASHKODEKDFPMSGVNIFGKE--SAGASNTALDNVITDEEIKATNPVATERFG 579
```

```
Db 516 PVATTTHTEDQASSVPAQNIIGIAKDPYRSGSTLAGIDIMVTDOEIAPTNGVGWRPYGL 575
QY 580 VAVNFOSSSTDPATGVDHVMGALPGMVWODRDVYLOGPIWAKIPIHTDGHFHPSPLMGGG 639
Db 576 TVTNEQNTTTAPTNAEVLGALPGMVWQNDIYLOGPIWAKIPIKTDGKFPSPNLGGG 635
QY 640 LKNPPPPQILIKNTVPANPPAPFSAATKFPASITQYSTQGVSVIEWELQKENSKRWNPEV 699
Db 636 LKNPPPPQVFIKNTVPADPPLEYNQKWSYITQYSTQGVTEMVWELURKENSKEWNPFI 695
QY 700 QYTSNYAKSANVDFTVNNGLYTPRPIGTRYLRL 736
Db 696 QYTSNYAKSANVDFTVNNGLYTPRPIGTRYLRL 732

RESULT 13
Q87666 PRELIMINARY; PRT; 732 AA.
AC Q87666;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE VPI.
GN VPI.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Virulent B;
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
duck parvoviruses indicates common ancestral origin with adeno-
associated virus 2.";
RL Virology 212:562-573 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Virulent B;
RA Zadori Z.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25749; AAA83230.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC64;

Query Match 54.5%; Score 2173.5; DB 12; Length 732;
Best Local Similarity 54.7%; Pred. No. 2.8e-141;
Matches 414; Conservative 108; Mismatches 176; Indels 59; Gaps 14;

QY 9 DWLEDNLSEGIREWMDLPGAPKPKANQKQ-----DDGRGLVLPQYKYLGP 57
Db 6 DSFEWYETAASWRHLKAGAPKPKNQSQSVSDREPERKDNNGRGFVLPQYKYLGP 65
QY 58 GLDKGPPVNAADAALHDKAYDQOLKAGDNPVLRYNHADAEFQERLQEDTSFGNGLGRA 117
Db 66 GLDKGPPVNAADSVALEHDKAYDQOLKAGDNPYIKFNHADQEFIDNLQDTSFGNGLGRA 125
QY 118 VFQAKRVLPLGLVEEGAKTAPGKR-----PVRSQPO-BPDSSSGIGTKTQGPAPK 168
Db 126 VFQAKRILPLGLVEEPVNTAPAKNTGKLTDTHTYPPVKKPKLTDYEVAGGSSAVQDG- 184
QY 169 KRLNFGQTDSESVDPDQPLGEPATPAAGVPTTMASGGGAPMADNNEGADGVGNAS 228
Db 195 -----GATAE-----GTEP-----VAASEMAEGGGAMGDSGAGGVGNAS 224
QY 229 HCDSTWLGDRVITSTRTWALPTNNHLYKQISSASTGASNDNHYFGYSTPWGYFDN 286
Db 225 HCDSTWLGDRVITSTRTWALPTNNHLYKQISSASTGASNDNHYFGYSTPWGYFDN 281
QY 287 RFHCHFSRDMQRLNNHNGFRPKRLNFKLNIQVKEVTNDGVTIANNLSTVQVFS 346
```

```

Db 282 RFHCFSPRQRLNNHNGIRPKSLKFKIFNVQVKEVTTDQTKTIANNLSTIQVFTD 341
Qy 347 SEYQLPVVLGSAHOGCLPPFPADVFMIPQYGLTLN---NGSOAVGRSSFCYLEYPPSQM 403
Db 342 DEHQLPVVLGSATGTMPPFPDVALPQYGYCTMHTNQNGARFNDRSAPFCYLEYPPSQM 401
Qy 404 LRTGNFTFSYTFEEVPHSSAHOSQSLDRMLNPLIDQYLYLNRTQNGSSAQNKLDF 463
Db 402 LRTGNFTFSYTFEEVPHSSAHOSQSLDRMLNPLIDQYLYLNRTQNGSSAQNKLDF 455
Qy 464 SRGSPAGMSVQPKWLPKPCYRQV---SKTKTNNNSNFT---WTGASKYNLNGRESIING 521
Db 456 KKAAGKAGTGMGRNWLPGKFLDQVRAYTGGTDNANMNIWSNGKNVLDKQYLLQPG 515
Qy 522 TAMASHKDDKDFPMSGVMIIFGKE--SAGASNTALDNVMTDEEBEIKATNPVATERFGT 579
Db 516 PVSATVTEGEASSLPQAQNLGIAKDPYRSGSTTAGISDIMVTEEQEVAPINGVWKPYGR 575
Qy 580 VAVNFQSSSTDATGDVHAMGALPGMWQDRDYVLOGPIWAKIPHTDGHFHPSPMLGGFG 639
Db 576 TVTNEQNTTAPTASDLDVLGALPGMWQNRDIYVLOGPIWAKIPKTDGKPHSPNLGGFG 635
Qy 640 LKNPPQIILKNTVPANPAEFSATKFASTFYQSTGVSVIEIEMWELKENSKRWNPEV 699
Db 636 LKNPPQVFIKNTVPADPPEVYHOKWNSIYQSTGQCTVEMWELKENSKRWNPEI 695
Qy 700 QYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
Db 696 QYTSNFSNRSTIMFAPNETGGYVEDRLIGTRYLTRPL 732

RESULT 14
Q83289 PRELIMINARY; PRT; 732 AA.
AC Q83289;
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DE 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE VPI protein.
OS Muscovy duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=37325;
RN [1]
RP SEQUENCE FROM N.A.
RA Zadori Z., Erdei J., Nagy J., Kleati J.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X75093; CAA52984.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81284 MW; 9FF6955BC15DF3AA CRC64;

Query Match 54.4%; Score 2168.5; DB 12; Length 732;
Best Local Similarity 55.2%; Pred. No. 6.1e-141;
Matches 418; Conservative 90; Mismatches 186; Indels 63; Gaps 13;

Qy 9 DWLEDLNLSGIREWDLKPGAPKPKANQKQKQ-----DDGRLVLPYKYLGPEN 57
Db 10 DWYE----TAAASWRHLKAGAPKPKNSQSQSVSTDRKPKQKDNNGRGVLPYKYLGPEN 65
Qy 58 GLDKGPPVNAADAALHDKAYDQOLKAGDNPYLRNHNADAEFQERLQEDTSFGNGLGRA 117
Db 66 GLDKGPPVNAKADVALEHDKAYDQOLKAGDNPYIKFNHADQDFIDSLQDDRSFGNGLGKA 125
Qy 118 VFOAKKRVLEPLGLVEGAKTAPKPKRPEVQSPQEPDSSSGIGTKTQQAQKRLNFGQTG 177
Db 126 VFOAKKRVLEPLGLVEPNTAPAK-----SSG-KLTDDHDFVKKPKLSE-- 170
Qy 178 DSSESVPDPQPLG-----EPPATPAVGTMTASGGGAPMADNNEGADGVGNASG 226
Db 171 --ENSPPSPNSGSGEASAAATEGSEPVAA-----NMAEGSGAMGDSAGGADGVGNASG 222
Qy 227 NWHCDSTLWLDGRVITSTRTWALPTNNHLYKQISSASTGASNDNHYFGYSTPWPWFYDEN 286

```

```

Db 223 NWHCDSTLWLDGRVITSTRTWALPSYNNHMYQAITSGTNPSN--TQYAGYSTPWPWFYDEN 281
Qy 287 RFHCFSPRQRLNNHNGIRPKSLKFKIFNVQVKEVTTDQTKTIANNLSTIQVFTD 346
Db 282 RFHCFSPRQRLNNHNGIRPKSLKFKIFNVQVKEVTTDQTKTIANNLSTIQVFTD 341
Qy 347 SEYQLPVVLGSAHOGCLPPFPADVFMIPQYGLTLN---NGSOAVGRSSFCYLEYPPSQM 403
Db 342 DEHQLPVVLGSATGTMPPFPDVALPQYGYCTMHTNQNGARFNDRSAPFCYLEYPPSQM 401
Qy 404 LRTGNFTFSYTFEEVPHSSAHOSQSLDRMLNPLIDQYLYLNRTQNGSSAQNKLDF 463
Db 402 LRTGNFTFSYTFEEVPHSSAHOSQSLDRMLNPLIDQYLYLNRTQNGSSAQNKLDF 455
Qy 464 SRGSPAGMSVQPKWLPKPCYRQV---SKTKTNNNSNFT---WTGASKYNLNGRESIING 521
Db 456 KKAAGKAGTGMGRNWLPGKFLDQVRAYTGGTDNANMNIWSNGKNVLDKQYLLQPG 515
Qy 522 TAMASHKDDKDFPMSGVMIIFGKE--SAGASNTALDNVMTDEEBEIKATNPVATERFGT 579
Db 516 PVSATVTEGEASSLPQAQNLGIAKDPYRSGSTTAGISDIMVTEEQEVAPINGVWKPYGR 575
Qy 580 VAVNFQSSSTDATGDVHAMGALPGMWQDRDYVLOGPIWAKIPHTDGHFHPSPMLGGFG 639
Db 576 TVTNEQNTTAPTASDLDVLGALPGMWQNRDIYVLOGPIWAKIPKTDGKPHSPNLGGFG 635
Qy 640 LKNPPQIILKNTVPANPAEFSATKFASTFYQSTGVSVIEIEMWELKENSKRWNPEV 699
Db 636 LKNPPQVFIKNTVPADPPEVYHOKWNSIYQSTGQCTVEMWELKENSKRWNPEI 695
Qy 700 QYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
Db 696 QYTSNFSNRSTIMFAPNETGGYVEDRLIGTRYLTRPL 732

RESULT 15
Q8V395 PRELIMINARY; PRT; 732 AA.
AC Q8V395;
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Capsid protein VP.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RA Ge Y., You Y., Xu Q.;
RL "Analysis of the major open reading frames' nucleotide sequences in
RT Goose parvovirus GPV-YG strain isolated in China.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF416726; AAL37722.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81456 MW; 73F2E4BC769744B6 CRC64;

Query Match 54.3%; Score 2165.5; DB 12; Length 732;
Best Local Similarity 54.8%; Pred. No. 9.9e-141;
Matches 415; Conservative 101; Mismatches 182; Indels 59; Gaps 14;

Qy 9 DWLEDLNLSGIREWDLKPGAPKPKANQKQKQ-----DDGRLVLPYKYLGPEN 57
Db 6 DSFEWEYETAASWRHLKAGAPKPKNSQSQSVSPAREPERKDSNRGVLPYKYLGPEN 65
Qy 58 GLDKGPPVNAADAALHDKAYDQOLKAGDNPYLRNHNADAEFQERLQEDTSFGNGLGRA 117
Db 66 GLDKGPPVNAKADVALEHDKAYDQOLKAGDNPYIKFNHADQDFIDSLQDDRSFGNGLGKA 125
Qy 118 VFOAKKRVLEPLGLVEGAKTAPKPKR-----PVQSPQ-EPDSSSGIGTKTQQAQK 168

```

Db 126 VFOAKKRILEPFLVEDSINTAPVKKNTGKLTLDHYVVKPKLTEEVSAGGTSVAVQDG- 184  
Qy 169 KRLNFGOTGDSVPDPPLGEPPTAPAVGPTTMSGGAPMADNNEGADGVGNASGNW 228  
Db 185 -----GATAE-----GTEP-----VAASEMAEGGALGDAGDGVGNASGNW 224  
Qy 229 HCDSTWLGDRVITTSRTWALPTYNHLYKOISSASTGASNDN--HYFGYSTPMGYFDFN 286  
Db 225 HCDSQMGNVTITKTRTWLPSYNNHIYKAITS---GTSQDATVQYAGYSTPMGYFDFN 281  
Qy 287 RFCHFSRDRMORLNNWGFPRKPLNFKNIOKEVTTNDGVTTIANNLTSTVQVPSD 346  
Db 282 RFCHFSRDRMORLNNWGFPRKPLNFKNIOKEVTTNDGVTTIANNLTSTVQVPSD 341  
Qy 347 SEYQLPYVLGSAHQCLPPFPADVEMIPQYGYLTJLN---NGSOAVGRSSFYCLEYFPPSQM 403  
Db 342 DEHQLPYVLGSAHQCLPPFPADVEMIPQYGYLTJLN---NGSOAVGRSSFYCLEYFPPSQM 401  
Qy 404 LRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQLYYLNRTONQSGSAQNKDLFF 463  
Db 402 LRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQLYYLNRTONQSGSAQNKDLFF 455  
Qy 464 SRGSPAGMSVQPKNMLPGCYTQORV-SKTKTDNNNSFT-WTGASKYNLNGRBSIINPG 521  
Db 456 KKA VKGAYGTMGRLNPLPGPKLLDQVRVAYTGCTDNYANWNIWSNGNKVNLKDRQYLLQPG 515  
Qy 522 TAMASHKDEDEKFFPMGSMWIFGKE--SAGASNTALDNVMTDEBEIKATNPVATERGT 579  
Db 516 PVSATHTKVEASSIPAQNLGLAKOPYRSGSTTAGISDIMVTDQEVAPTNGVGWPKYK 575  
Qy 580 VAVNFQSSSTDPAIGDVHAMGALPGMWQDRDYLQGPWAKIPHTDGHFHPSPMLGGFG 639  
Db 576 TVTNEQNTTTAPTSSDLVDLGLPGMWQDRDYLQGPWAKIPHTDGHFHPSPMLGGFG 635  
Qy 640 LKNPPPOILLKNTVPANPPAFSATKPAFITYSTQGVSVYEWELOKENSKENWNEV 699  
Db 636 LYNPPPOVFIKNTVPANPPAFSATKPAFITYSTQGVSVYEWELOKENSKENWNEV 695  
Qy 700 QYTSNYAKSANYDFTVDNNGLYTEPRPIGTTRYLTP 736  
Db 696 QYTSNFSDRTSINFAPNETGGYIEDRLIGTRYLTP 732

Search completed: January 21, 2004, 16:06:09  
Job time : 42.198 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:04:27 ; Search time 33.0787 Seconds  
(without alignments)  
4549.706 Million cell updates/sec

Title: US-09-807-802a-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWLENLSEIGIR.....NNGLYTEPRIGTYLTRL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3989	100.0	736	12	US-10-291-583-64
2	3989	100.0	736	12	US-10-423-704A-5
3	3983	99.3	736	12	US-10-291-583-65
4	3494	87.6	736	12	US-10-291-583-71
5	3494	87.6	736	12	US-10-423-704A-6
6	3473	87.1	736	12	US-10-291-583-97
7	3467	86.9	736	12	US-10-291-583-100
8	3467	86.9	736	12	US-10-423-704A-7
9	3464	86.8	736	12	US-10-291-583-96
10	3460	86.7	736	12	US-10-291-583-98
11	3457	86.7	736	12	US-10-291-583-99
12	3453	86.6	738	12	US-10-291-583-85
13	3448	86.4	738	12	US-10-291-583-79
14	3448	86.4	738	12	US-10-291-583-81
15	3445	86.4	738	12	US-10-291-583-93

16	3445	86.4	738	12	US-10-291-583-94
17	3442.5	86.3	737	12	US-10-291-583-2
18	3442.5	86.3	737	12	US-10-423-704A-8
19	3441	86.3	738	12	US-10-291-583-80
20	3440	86.2	738	12	US-10-291-583-84
21	3438	86.2	738	12	US-10-291-583-92
22	3434	86.1	728	12	US-10-291-583-108
23	3432	86.0	738	12	US-10-291-583-91
24	3429	86.0	738	12	US-10-291-583-82
25	3424.5	85.8	737	12	US-10-291-583-72
26	3422.5	85.8	729	12	US-10-291-583-110
27	3422	85.8	738	12	US-10-291-583-83
28	3420.5	85.7	733	12	US-10-291-583-86
29	3420	85.7	728	12	US-10-291-583-102
30	3417.5	85.7	729	12	US-10-291-583-111
31	3416.5	85.6	733	12	US-10-291-583-87
32	3415.5	85.6	735	12	US-10-291-583-67
33	3415	85.6	728	12	US-10-291-583-106
34	3414.5	85.6	729	12	US-10-291-583-109
35	3414.5	85.6	731	12	US-10-291-583-89
36	3413.5	85.6	735	12	US-10-291-583-68
37	3413.5	85.6	735	12	US-10-291-583-112
38	3413	85.6	728	12	US-10-291-583-107
39	3411.5	85.5	735	12	US-10-291-583-69
40	3408.5	85.4	731	12	US-10-291-583-88
41	3407.5	85.4	735	12	US-10-291-583-66
42	3406	85.4	728	12	US-10-291-583-105
43	3405	85.4	738	12	US-10-291-583-95
44	3405	85.4	738	12	US-10-423-704A-2
45	3402.5	85.3	735	12	US-10-291-583-70

ALIGNMENTS

RESULT 1

US-10-291-583-64  
; Sequence 64, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (US20030138772A1) Sequences Identifi  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64  
; LENGTH: 736  
; TYPE: PRT

US-10-291-583-64  
; ORGANISM: capsid protein of AAV serotype, clone AAV1

Query Match	100.0%	Score	3989;	DB	12;	Length	736;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	736;	Conservative	0;	Mismatches	0;	Gaps	0;
Oy	1	MAADGYLPDWLENLSEIGIREWDLKPGAPKPKANQKQDDGRGLVLPGYKYLGPFGNGLD	60				
Db	1	MAADGYLPDWLENLSEIGIREWDLKPGAPKPKANQKQDDGRGLVLPGYKYLGPFGNGLD	60				
Oy	61	KGEPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEPQERLQEDTSFGGNLGRAVPQ	120				



```
Db 61 KGPVNAADAAALEHDKAYDQQLKAGDNPLYRYNHADAEFQERLQEDTSFGNLRGAVFQ 120
Qy 121 AKKRVLEPLGLVEGAKTAPGKRPVQSPQSDSSGIGTKGQOPAKKRLNFGOTGSE 180
Db 121 AKKRVLEPLGLVEGAKTAPGKRPVQSPQSDSSGIGTKGQOPAKKRLNFGOTGSE 180
Qy 181 SVPDPQPLGEPATPAAVGPPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240
Db 181 SVPDPQPLGEPATPAAVGPPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240
Qy 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPFDRNRFCHFSPRDWORL 300
Db 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPFDRNRFCHFSPRDWORL 300
Qy 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTTIANNLSTVQFSDSEYQLPYVLGSAHQ 360
Db 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTTIANNLSTVQFSDSEYQLPYVLGSAHQ 360
Qy 361 GCLPPPPADVFMIPOYGYLTLNNGSQAAGRSSFYCLEYFPQSOMLRTGNNFTFSYTFEEVP 420
Db 361 GCLPPPPADVFMIPOYGYLTLNNGSQAAGRSSFYCLEYFPQSOMLRTGNNFTFSYTFEEVP 420
Qy 421 FHSSYAHQSQSLDRLMNPLIDQLYLNLRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWL 480
Db 421 FHSSYAHQSQSLDRLMNPLIDQLYLNLRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWL 480
Qy 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLAGRESIINPGTAMASHKDDKDFPMSGV 540
Db 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLAGRESIINPGTAMASHKDDKDFPMSGV 540
Qy 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Db 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Qy 601 ALPGMWQDRDYLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPANPPA 660
Db 601 ALPGMWQDRDYLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPANPPA 660
Qy 661 EFSATKFASTFYQYSTGQVSVEIEWELQKENSQRWNPVEQVTSNYAKSANVDFTVDNNGL 720
Db 661 EFSATKFASTFYQYSTGQVSVEIEWELQKENSQRWNPVEQVTSNYAKSANVDFTVDNNGL 720
Qy 721 YTEPRPIGTRYLTRPL 736
Db 721 YTEPRPIGTRYLTRPL 736

RESULT 2
US-10-423-704A-5
; Sequence 5, Application US/10423704A
; Publication No. US20030228282A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: Adeno-Associated Virus (AAV) Serotype 8 Sequences, Vectors
; TITLE OF INVENTION: Containing Same, and Uses Therefor
; FILE REFERENCE: UPN-02733AUSA
; CURRENT APPLICATION NUMBER: US/10/423,704A
; PRIOR FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: US 60/341,151
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,133
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,122
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: PCT/US02/33630
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 736
```

```
; TYPE: PRT
; ORGANISM: adeno-associated virus serotype 1
US-10-423-704A-5

Query Match 100.0%; Score 3989; DB 12; Length 736;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAADGYLPDWLEEDNLSEGIREWMDLKPQAPKPKANQKQDDGRGLVLPGYKYLGPFGNGLD 60
Db 1 MAADGYLPDWLEEDNLSEGIREWMDLKPQAPKPKANQKQDDGRGLVLPGYKYLGPFGNGLD 60
Qy 61 KGPVNAADAAALEHDKAYDQQLKAGDNPLYRYNHADAEFQERLQEDTSFGNLRGAVFQ 120
Db 61 KGPVNAADAAALEHDKAYDQQLKAGDNPLYRYNHADAEFQERLQEDTSFGNLRGAVFQ 120
Qy 121 AKKRVLEPLGLVEGAKTAPGKRPVQSPQSDSSGIGTKGQOPAKKRLNFGOTGSE 180
Db 121 AKKRVLEPLGLVEGAKTAPGKRPVQSPQSDSSGIGTKGQOPAKKRLNFGOTGSE 180
Qy 181 SVPDPQPLGEPATPAAVGPPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240
Db 181 SVPDPQPLGEPATPAAVGPPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240
Qy 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPFDRNRFCHFSPRDWORL 300
Db 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPFDRNRFCHFSPRDWORL 300
Qy 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTTIANNLSTVQFSDSEYQLPYVLGSAHQ 360
Db 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTTIANNLSTVQFSDSEYQLPYVLGSAHQ 360
Qy 361 GCLPPPPADVFMIPOYGYLTLNNGSQAAGRSSFYCLEYFPQSOMLRTGNNFTFSYTFEEVP 420
Db 361 GCLPPPPADVFMIPOYGYLTLNNGSQAAGRSSFYCLEYFPQSOMLRTGNNFTFSYTFEEVP 420
Qy 421 FHSSYAHQSQSLDRLMNPLIDQLYLNLRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWL 480
Db 421 FHSSYAHQSQSLDRLMNPLIDQLYLNLRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWL 480
Qy 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLAGRESIINPGTAMASHKDDKDFPMSGV 540
Db 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLAGRESIINPGTAMASHKDDKDFPMSGV 540
Qy 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Db 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Qy 601 ALPGMWQDRDYLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPANPPA 660
Db 601 ALPGMWQDRDYLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPANPPA 660
Qy 661 EFSATKFASTFYQYSTGQVSVEIEWELQKENSQRWNPVEQVTSNYAKSANVDFTVDNNGL 720
Db 661 EFSATKFASTFYQYSTGQVSVEIEWELQKENSQRWNPVEQVTSNYAKSANVDFTVDNNGL 720
Qy 721 YTEPRPIGTRYLTRPL 736
Db 721 YTEPRPIGTRYLTRPL 736

RESULT 3
US-10-291-583-65
; Sequence 65, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1e1 Sequences Identifi
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
```

; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 65  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone AAV6VPI  
US-10-291-583-65

Query Match 99.3%; Score 3963; DB 12; Length 736;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 730; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAADGYPDLWLENTSEGIREWDLKPGAPKPKANQKQDDGRGLVPGYKYLGPENGLD 60  
DB 1 MAADGYPDLWLENTSEGIREWDLKPGAPKPKANQKQDDGRGLVPGYKYLGPENGLD 60

QY 61 KGEVPVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGNLRGAVFQ 120  
DB 61 KGEVPVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGNLRGAVFQ 120

QY 121 AKRVLEPLGLVEGAKTAPGKRPEQSPQEPDSSSGIGTKQQQPAKRLNFGQTDSE 180  
DB 121 AKRVLEPLGLVEGAKTAPGKRPEQSPQEPDSSSGIGTKQQQPAKRLNFGQTDSE 180

QY 181 SVDPDPLGEPBPATPAAGVPTTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
DB 181 SVDPDPLGEPBPATPAAGVPTTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240

QY 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRL 300  
DB 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRL 300

QY 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQ 360  
DB 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQ 360

QY 361 GCLPPPPADVPMIPOYGYLTLLNGSQAVGRSSFCYCLEYPPSQMLRTGNNFTSYTFEVP 420  
DB 361 GCLPPPPADVPMIPOYGYLTLLNGSQAVGRSSFCYCLEYPPSQMLRTGNNFTSYTFEVP 420

QY 421 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQNSGSAQNKDLLFSRGS PAGMSVQPKNWL 480  
DB 421 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQNSGSAQNKDLLFSRGS PAGMSVQPKNWL 480

QY 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPFMSGV 540  
DB 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPFMSGV 540

QY 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMG 600  
DB 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMG 600

QY 601 ALPGVMWQDRDYYLQGPFIWAKIPHTDGHFHPSPLMGGFGLKXNPPPIKNTVPVNPAPA 660  
DB 601 ALPGVMWQDRDYYLQGPFIWAKIPHTDGHFHPSPLMGGFGLKXNPPPIKNTVPVNPAPA 660

QY 661 EFSATKFASTIQTQYSTGQVSVEIWELOKENS KRWNPVEQVYTSNYAKSANVDTVDNGL 720  
DB 661 EFSATKFASTIQTQYSTGQVSVEIWELOKENS KRWNPVEQVYTSNYAKSANVDTVDNGL 720

QY 721 YTEPRPIGTRYLTRPL 736  
DB 721 YTEPRPIGTRYLTRPL 736

RESULT 4  
US-10-291-583-71  
; Sequence 71, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (  
; FILE REFERENCE: Sequences and Isolating No. US20030138772A1el Sequences Identifi  
; CURRENT APPLICATION NUMBER: US/10/291.583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 71  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone AAV3  
US-10-291-583-71

Query Match 87.6%; Score 3494; DB 12; Length 736;  
Best Local Similarity 86.4%; Pred. No. 2.2e-295;  
Matches 637; Conservative 39; Mismatches 59; Indels 2; Gaps 2;

QY 1 MAADGYPDLWLENTSEGIREWDLKPGAPKPKANQKQDDGRGLVPGYKYLGPENGLD 60  
DB 1 MAADGYPDLWLENTSEGIREWDLKPGAPKPKANQKQDDGRGLVPGYKYLGPENGLD 60

QY 61 KGEVPVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGNLRGAVFQ 120  
DB 61 KGEVPVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGNLRGAVFQ 120

QY 121 AKRVLEPLGLVEGAKTAPGKRPEQSPQEPDSSSGIGTKQQQPAKRLNFGQTDSE 180  
DB 121 AKRVLEPLGLVEGAKTAPGKRPEQSPQEPDSSSGIGTKQQQPAKRLNFGQTDSE 180

QY 181 SVDPDPLGEPBPATPAAGVPTTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
DB 181 SVDPDPLGEPBPATPAAGVPTTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240

QY 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRL 300  
DB 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRL 300

QY 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQ 360  
DB 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQ 360

QY 361 GCLPPPPADVPMIPOYGYLTLLNGSQAVGRSSFCYCLEYPPSQMLRTGNNFTSYTFEVP 420  
DB 361 GCLPPPPADVPMIPOYGYLTLLNGSQAVGRSSFCYCLEYPPSQMLRTGNNFTSYTFEVP 420

QY 421 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQNSGSAQNKDLLFSRGS PAGMSVQPKNWL 479  
DB 421 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQNSGSAQNKDLLFSRGS PAGMSVQPKNWL 479

QY 480 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPFMSGV 539  
DB 480 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPFMSGV 539

QY 540 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAM 599  
DB 540 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAM 599



Db 181 SVDPDPLGPPAPSPGLGNTWASGGCAPMADNNEGADGVGSSGNHCHDSTWLGDRVI 240  
Qy 241 TTSTRTWALPTNNHLYKQISSASTGAS - NDNHVFGYSTPWGYFDNRFRCHFSPRDWOR 299  
Db 241 TTSTRTWALPTNNHLYKQISSASTGAS - NDNHVFGYSTPWGYFDNRFRCHFSPRDWOR 300  
Qy 300 LINNMWGFPRKLNFKLFIQVKEVTNDGVTTIANNLSTVQVFDSEYQLPYVLGSAH 359  
Db 301 LINNMWGFPRKLNFKLFIQVKEVTNDGVTTIANNLSTVQVFDSEYQLPYVLGSAH 360  
Qy 360 QGCLPPFPADVPMI PQGYLTLNNGSQAVGRSSFYCLEYFPSSQMLRTGNFTSYTFEEV 419  
Db 361 QGCLPPFPADVPMI PQGYLTLNNGSQAVGRSSFYCLEYFPSSQMLRTGNFTSYTFEEV 420  
Qy 420 PFHSSYAHQSOLDRMLNPLIDQYLYLNRTONOSGAQNKDLLFSRGSAGMSVQPKNWL 479  
Db 421 PFHSSYAHQSOLDRMLNPLIDQYLYLNRTONOSGAQNKDLLFSRGSAGMSVQPKNWL 479  
Qy 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSG 539  
Db 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSG 539  
Qy 540 VMIFGKESAGASNTALDNVMTDEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAM 599  
Db 540 VLIQKQAGNDGVYDYSQVLTDEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAM 599  
Qy 600 GALPGMWQDRDVLQGPWAKI PHTDGHFHPSPLMGGFGLKNNPPPOILLIKNTVPANPP 659  
Db 600 GVIPGMWQNRDVLQGPWAKI PHTDGNFHPSPLMGGFGLKNNPPPOILLIKNTVPANPP 659  
Qy 660 AEFSAFKFASFITQYSTQGVSVIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 719  
Db 660 LTFNQAKLSNFITQYSTQGVSVIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 719  
Qy 720 LYTEPRPIGTRYLTRPL 736  
Db 720 VYSEPRPIGTRYLTRNL 736

RESULT 7  
US-10-291-583-100  
; Sequence 100, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 100  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone AAV9  
US-10-291-583-100

Query Match 86.9%; Score 3467; DB 12; Length 736;  
Best Local Similarity 85.5%; Pred. No. 4.9e-293;  
Matches 630; Conservative 44; Mismatches 61; Indels 2; Gaps 2;

Qy 1 MAADGYLPDLWLDNLSEGIREWMDLKPAPKPKANQKQDQDGRGLVLPGYKYGLPPNGLD 60  
Db 1 MAADGYLPDLWLDNLSEGIREWMDLKPAPKPKANQKQDQDGRGLVLPGYKYGLPPNGLD 60  
Qy 61 KGEPPVNAADAALAEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQBDTSFGNGLGRAVQ 120  
Db 61 KGEPPVNAADAALAEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQBDTSFGNGLGRAVQ 120  
Qy 121 AKKRVLEPLGLVBEAGAKTAPGKKRPVBOQSPQDSSSGIKGTGQOQAKKRLNFGQTDSE 180  
Db 121 AKKRVLEPLGLVBEAGAKTAPGKKRPVBOQSPQDSSSGIKGTGQOQAKKRLNFGQTDSE 180  
Qy 181 SVDPDPLGPPAPSPGLGNTWASGGCAPMADNNEGADGVGSSGNHCHDSTWLGDRVI 240  
Db 181 SVDPDPLGPPAPSPGLGNTWASGGCAPMADNNEGADGVGSSGNHCHDSTWLGDRVI 240  
Qy 241 TTSTRTWALPTNNHLYKQISSASTGAS - NDNHVFGYSTPWGYFDNRFRCHFSPRDWOR 299  
Db 241 TTSTRTWALPTNNHLYKQISSASTGAS - NDNHVFGYSTPWGYFDNRFRCHFSPRDWOR 300  
Qy 300 LINNMWGFPRKLNFKLFIQVKEVTNDGVTTIANNLSTVQVFDSEYQLPYVLGSAH 359  
Db 301 LINNMWGFPRKLNFKLFIQVKEVTNDGVTTIANNLSTVQVFDSEYQLPYVLGSAH 360  
Qy 360 QGCLPPFPADVPMI PQGYLTLNNGSQAVGRSSFYCLEYFPSSQMLRTGNFTSYTFEEV 419  
Db 361 QGCLPPFPADVPMI PQGYLTLNNGSQAVGRSSFYCLEYFPSSQMLRTGNFTSYTFEEV 420  
Qy 420 PFHSSYAHQSOLDRMLNPLIDQYLYLNRTONOSGAQNKDLLFSRGSAGMSVQPKNWL 479  
Db 421 PFHSSYAHQSOLDRMLNPLIDQYLYLNRTONOSGAQNKDLLFSRGSAGMSVQPKNWL 479  
Qy 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSG 539  
Db 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSG 539  
Qy 540 VMIFGKESAGASNTALDNVMTDEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAM 599  
Db 540 VLIQKQAGNDGVYDYSQVLTDEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAM 599  
Qy 600 GALPGMWQDRDVLQGPWAKI PHTDGHFHPSPLMGGFGLKNNPPPOILLIKNTVPANPP 659  
Db 600 GVIPGMWQNRDVLQGPWAKI PHTDGNFHPSPLMGGFGLKNNPPPOILLIKNTVPANPP 659  
Qy 660 AEFSAFKFASFITQYSTQGVSVIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 719  
Db 660 LTFNQAKLSNFITQYSTQGVSVIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 719  
Qy 720 LYTEPRPIGTRYLTRPL 736  
Db 720 VYSEPRPIGTRYLTRNL 736

RESULT 8  
US-10-423-704A-7  
; Sequence 7, Application US/10423704A  
; Publication No. US20030228282A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: Adeno-Associated Virus (AAV) Serotype 8 Sequences, Vectors  
; FILE REFERENCE: UPN-02733AUSA  
; CURRENT APPLICATION NUMBER: US/10/423,704A  
; CURRENT FILING DATE: 2003-04-25  
; PRIOR APPLICATION NUMBER: US 60/341,151  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,133  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,122  
; PRIOR FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: PCT/US02/33630

; PRIOR FILING DATE: 2002-11-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: adeno-associated virus serotype 9  
US-10-423-704A-7

```
Query Match      86.9%; Score 3467; DB 12; Length 736;
Best Local Similarity 85.5%; Pred. No. 4.9e-293;
Matches 630; Conservative 44; Mismatches 61; Indels 2; Gaps 2;

QY 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQQKDDGRGLVLPYKYKLGPFNGLD 60
DB 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQQKDDGRGLVLPYKYKLGPFNGLD 60

QY 61 KGPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRGAVPQ 120
DB 61 KGPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRGAVPQ 120

QY 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQDPDSSGGIGKTGQOPAKKRLNFGOTGDS 180
DB 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQDPDSSGGIGKTGQOPAKKRLNFGOTGDS 180

QY 181 SVDPDQPLGEPAPTAAGVPTTMSAGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVI 240
DB 181 SVDPDQPLGEPAPTAAGVPTTMSAGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVI 240

QY 241 TTSTRTWALPTYNNHLYKQISASSTGAS -NDNHFGYSTPMGYDFNRFCHFSPRDQW 299
DB 241 TTSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGYSTPMGYDFNRFCHFSPRDQW 300

QY 300 LINNNWGFPRKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAH 359
DB 301 LINNNWGFPRKRLNFKLFNIQVKEVTTNEGTKTIANNLTSTVRVFTDSEYQLPYVLGSAH 360

QY 360 QGCLPPPPADVFMIPOGYLYTLNNGSQAVGRSSFYCLEYFPFSQMLRTGNFTFSYTFEEV 419
DB 361 QGCLPPPPADVFVMPQGYLYTLNNGSQALGRSSFYCLEYFPFSQMLRTGNFTFSYTFEDV 420

QY 420 PHSSYAHQSQSLDRMLNPLIDQYLYLNRTOQSGSAQNKDLLFSRGSAGMSVOPQKWL 479
DB 421 PHSSYAHQSQSLDRMLNPLIDQYLYLVRTQT -TGTTGTTQLAFSQAGPSSMANQARNV 479

QY 480 PGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSG 539
DB 480 PGPCYRQQRVSTTTNQNNNSNFAWTGAAKFKLNGRDSLNNPGVAMASHKDDKDFPSSG 539

QY 540 VMIFGKESAGSNTALDNVMTDDEEIKATNPVATERFGTVAVNFOSSTDPATGDVHAM 599
DB 540 VLIFFGQAGNDGVDSQVLIITDDEEIKATNPVATEYGAIVANNQAANTOAGTGLVHNQ 599

QY 600 GALPGVMWQDRDVLQGPWAKIPHTDGHFHPSPLMGGFGLKNNPPQIILIKTNPVPANPP 659
DB 600 GVTPGMWQNRDVLQGPWAKIPHTDGNFHPSPLMGGFGLKHPHPQIILIKTNPVPADPP 659

QY 660 AEFSATKFAFITQYSTGQVSVEIEWELQKENS KRNWPEVQYTSNYAKSANVDFTVDNNG 719
DB 720 LYTEPRPIGTRYLTRPL 736
DB 720 VYSEPRPIGTRYLTRNL 736
```

## RESULT 9

US-10-291-583-96  
; Sequence 96, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.

; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (f  
; FILE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identific  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 96  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 43.21  
US-10-291-583-96

Query Match 86.8%; Score 3464; DB 12; Length 736;  
Best Local Similarity 85.3%; Pred. No. 8.9e-293;  
Matches 629; Conservative 46; Mismatches 60; Indels 2; Gaps 2;

```
QY 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQQKDDGRGLVLPYKYKLGPFNGLD 60
DB 1 MAADGYLPDWLEDNLSGIREWDLKPGAPKPKANQQKDDGRGLVLPYKYKLGPFNGLD 60

QY 61 KGPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRGAVPQ 120
DB 61 KGPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRGAVPQ 120

QY 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQDPDSSGGIGKTGQOPAKKRLNFGOTGDS 180
DB 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQDPDSSGGIGKTGQOPAKKRLNFGOTGDS 180

QY 181 SVDPDQPLGEPAPTAAGVPTTMSAGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVI 240
DB 181 SVDPDQPLGEPAPTAAGVPTTMSAGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVI 240

QY 241 TTSTRTWALPTYNNHLYKQISASSTGAS -NDNHFGYSTPMGYDFNRFCHFSPRDQW 299
DB 241 TTSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGYSTPMGYDFNRFCHFSPRDQW 300

QY 300 LINNNWGFPRKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAH 359
DB 301 LINNNWGFPRKRLNFKLFNIQVKEVTTNEGTKTIANNLTSTVRVFTDSEYQLPYVLGSAH 360

QY 360 QGCLPPPPADVFMIPOGYLYTLNNGSQAVGRSSFYCLEYFPFSQMLRTGNFTFSYTFEEV 419
DB 361 QGCLPPPPADVFVMPQGYLYTLNNGSQALGRSSFYCLEYFPFSQMLRTGNFTFSYTFEDV 420

QY 420 PHSSYAHQSQSLDRMLNPLIDQYLYLNRTOQSGSAQNKDLLFSRGSAGMSVOPQKWL 479
DB 421 PHSSYAHQSQSLDRMLNPLIDQYLYLVRTQT -TGTTGTTQLAFSQAGPSSMANQARNV 479

QY 480 PGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSG 539
DB 480 PGPCYRQQRVSTTTNQNNNSNFAWTGAAKFKLNGRDSLNNPGVAMASHKDDKDFPSSG 539

QY 540 VMIFGKESAGSNTALDNVMTDDEEIKATNPVATERFGTVAVNFOSSTDPATGDVHAM 599
DB 540 VLIFFGQAGNDGVDSQVLIITDDEEIKATNPVATEYGAIVANNQAANTOAGTGLVHNQ 599

QY 600 GALPGVMWQDRDVLQGPWAKIPHTDGHFHPSPLMGGFGLKNNPPQIILIKTNPVPANPP 659
DB 600 GVTPGMWQNRDVLQGPWAKIPHTDGNFHPSPLMGGFGLKHPHPQIILIKTNPVPADPP 659

QY 660 AEFSATKFAFITQYSTGQVSVEIEWELQKENS KRNWPEVQYTSNYAKSANVDFTVDNNG 719
DB 660 LTFNQAKLSNFITQYSTGQVSVEIEWELQKENS KRNWPEIQTYSNYKSTNVDFAVNTEG 719
```

QY 720 LYTEPRPIGTRYLTRPL 736  
Db :|||||  
720 VYSEPRPIGTRYLTRNL 736  
Db :|||||  
RESULT 10  
US-10-291-583-98  
; Sequence 98, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (AAV) Sequences and Isolating No. US20030138772A1el Sequences Identifi  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 98  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 43.23  
US-10-291-583-98  
Query Match 86.7%; Score 3460; DB 12; Length 736;  
Best Local Similarity 85.3%; Pred. No. 2e-292;  
Matches 629; Conservative 44; Mismatches 62; Indels 2; Gaps 2;  
QY 1 MAADGYLPDWLENDLSGIREWMDLKPAPKPKANQKQDDGRGLVLPGYKYLGPNGLD 60  
Db 1 MAADGYLPDWLENDLSGIREWMDLKPAPKPKANQKQDDGRGLVLPGYKYLGPNGLD 60  
QY 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAEPQERLQEDTSFGNGLGRAVFQ 120  
Db 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAEPQERLQEDTSFGNGLGRAVFQ 120  
QY 121 AKRVLEPLGLVEGAKTAPGKRPVQSPQEPDSSSGIGTKTQQAQKRLNFGQTDSE 180  
Db 121 AKRVLEPLGLVEGAKTAPGKRPVQSPQEPDSSSGIGTKTQQAQKRLNFGQTDSE 180  
QY 181 SVDPQPLGPPPPATPAAGVPTTASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
Db 181 SVDPQPLGPPPPATPAAGVPTTASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
QY 241 TTSTRWTALPTYNHLYKQISSASTGAS - NDNHFGYSTPWGYFDNFRPHCFSPRDWQ 299  
Db 241 TTSTRWTALPTYNHLYKQISSASTGAS - NDNHFGYSTPWGYFDNFRPHCFSPRDWQ 300  
QY 300 LINNMGFRKRLNFKLFNIQVKEVTNDGVTIANNLTSTVQVFDSEYQLPYVLSAH 359  
Db 301 LINNMGFRKRLNFKLFNIQVKEVTNDGVTIANNLTSTVQVFDSEYQLPYVLSAH 360  
QY 360 QGCLPPPPADVPMIPOGYLTLLNNGSOAVGRSFCLEYPPSOMLRTGNFTSYTFEEV 419  
Db 361 QGCLPPPPADVPMIPOGYLTLLNNGSOAVGRSFCLEYPPSOMLRTGNFTSYTFEEV 420  
QY 420 PFHSSVAHSQSLDRMLNPLIDQYLYLNRNTQNSGSAQNKDLFSRGSAGMSVQPKNWL 479  
Db 421 PFHSSVAHSQSLDRMLNPLIDQYLYLNRNTQNSGSAQNKDLFSRGSAGMSVQPKNWL 479  
QY 480 PGPCYRQORVSKTKTDNNNSNFMTGAAGFKLNGRDSLMPGAMASHKDDDDRPFPSSG 539  
Db :|||||

Db 480 PGPCYRQORVSKTKTDNNNSNFMTGAAGFKLNGRDSLMPGAMASHKDDDDRPFPSSG 539  
QY 540 VMIKESAGASNTALDNVMTDEEIKATNPVATERFCTVAVNFQSSSTDDATGCVHAM 599  
Db :|||||  
540 VLIKQAGAGNDVYSQVLIITDEEIKATNPVATERFCTVAVNFQSSSTDDATGCVHAM 599  
QY 600 GALPGVMQDRDVLQGPITWAKIPHTDGHFHPSPLMGGFGLKNPPPIILIKNTVPANPP 659  
Db :|||||  
600 GVIPGMVQNRDVLQGPITWAKIPHTDGHFHPSPLMGGFGLKNPPPIILIKNTVPANPP 659  
QY 660 AFSATKFAFITQYSTQSVSEIWELOKENSKWNEVQVTSYAKSANYDFTVNNNG 719  
Db :|||||  
660 LTFNOAKLNSFITQYSTQSVSEIWELOKENSKWNEVQVTSYAKSANYDFTVNNNG 719  
QY 720 LYTEPRPIGTRYLTRPL 736  
Db :|||||  
720 VYSEPRPIGTRYLTRNL 736  
Db :|||||  
RESULT 11  
US-10-291-583-99  
; Sequence 99, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (AAV) Sequences and Isolating No. US20030138772A1el Sequences Identifi  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 99  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 43.20  
US-10-291-583-99  
Query Match 86.7%; Score 3457; DB 12; Length 736;  
Best Local Similarity 85.3%; Pred. No. 3.6e-292;  
Matches 629; Conservative 44; Mismatches 62; Indels 2; Gaps 2;  
QY 1 MAADGYLPDWLENDLSGIREWMDLKPAPKPKANQKQDDGRGLVLPGYKYLGPNGLD 60  
Db 1 MAADGYLPDWLENDLSGIREWMDLKPAPKPKANQKQDDGRGLVLPGYKYLGPNGLD 60  
QY 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAEPQERLQEDTSFGNGLGRAVFQ 120  
Db 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAEPQERLQEDTSFGNGLGRAVFQ 120  
QY 121 AKRVLEPLGLVEGAKTAPGKRPVQSPQEPDSSSGIGTKTQQAQKRLNFGQTDSE 180  
Db 121 AKRVLEPLGLVEGAKTAPGKRPVQSPQEPDSSSGIGTKTQQAQKRLNFGQTDSE 180  
QY 181 SVDPQPLGPPPPATPAAGVPTTASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
Db 181 SVDPQPLGPPPPATPAAGVPTTASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
QY 241 TTSTRWTALPTYNHLYKQISSASTGAS - NDNHFGYSTPWGYFDNFRPHCFSPRDWQ 299  
Db 241 TTSTRWTALPTYNHLYKQISSASTGAS - NDNHFGYSTPWGYFDNFRPHCFSPRDWQ 300  
QY 300 LINNMGFRKRLNFKLFNIQVKEVTNDGVTIANNLTSTVQVFDSEYQLPYVLSAH 359  
Db :|||||

Db 301 LINNNWGPRLNFKLNIQVKEVTNNEGKTIANNLTSTVQVFTDSEYQLPYVLGSAH 360  
Qy 360 OGCLPPPPADVFMIPOYGYLTNNGSQAVGRSSFCYCLEYFPSPQMLRTGNNTFESYTFEEV 419  
Db 361 OGCLPPPPADVFMIPOYGYLTNNGSQALGRSSFCYCLEYFPSPQMLRTGNNTFESYTFEEV 420  
Qy 420 PFHSSYAHSSQSLDLRLMNLIDQYLYLNRTOQSGSAQNKDLFLSRGSPAGMSVQPKNW 479  
Db 421 PFHSSYAHSSQSLDLRLMNLIDQYLYLNRTOQSGSAQNKDLFLSRGSPAGMSVQPKNW 479  
Qy 480 POPCYRQORVSKTKTDNNNSNFTWTCASKYNLNGRESIINPGTAMASHKODEDKFPFMSG 539  
Db 480 POPCYRQORVSKTKTDNNNSNFTWTCASKYNLNGRESIINPGTAMASHKODEDKFPFMSG 539  
Qy 540 VMI FGKESAGASNTALDNLVMTDEBEIKATNPVATERFGTVAVNFPQSSSTDPATGDVHAM 599  
Db 540 VLI FGKQAGNDVDYSQVLTDEBEIKATNPVATERFGTVAVNFPQSSSTDPATGDVHAM 599  
Qy 600 GALPGMWQDRDYLQGPWAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKNTVPVPANPP 659  
Db 600 GVI PGMWQNRDYLQGPWAKIPHTDGNFHPSPMLMGFGGLKNPPQILIKNTVPVPADPP 659  
Qy 660 AEFSAKTFASFTQYSTQGVSVIEWELOKENSKRNWPEVQYTSNYSKANSANTVDFTDNG 719  
Db 660 LTFNOAKLSNFTQYSTQGVSVIEWELOKENSKRNWPEVQYTSNYSKANSANTVDFAVNT 719  
Qy 720 LYTEPRPIGTRYLTRPL 736  
Db 720 VYSEPRPIGTRYLTRNL 736

RESULT 12  
US-10-291-583-85  
; Sequence 85, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; PRIOR FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 85  
; LENGTH: 738  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 42.8  
US-10-291-583-85

Query Match 86.8%; Score 3453; DB 12; Length 738;  
Best Local Similarity 85.1%; Pred. No. 8.1e-292;  
Matches 628; Conservative 45; Mismatches 63; Indels 2; Gaps 2;  
Qy 1 MAADGVLPLWLENTSEGIREWDLKPGAPKPKANQKQDDGRLVLPKYKYLGPFGNGLD 60  
Db 1 MAADGVLPLWLENTSEGIREWDLKPGAPKPKANQKQDDGRLVLPKYKYLGPFGNGLD 60  
Qy 61 KGEPPVNAADAALHDKAYDQQLKAGDNPYLRNTHADAEFQERLQEDTSGGNLGRAVFQ 120  
Db 61 KGEPPVNAADAALHDKAYDQQLKAGDNPYLRNTHADAEFQERLQEDTSGGNLGRAVFQ 120

Qy 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQ-BPDSSSGIGKGTGQOPAKRLNFGQTGDS 179  
Db 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQ-BPDSSSGIGKGTGQOPAKRLNFGQTGDS 180  
Qy 180 ESYVDDPOPLGEPPATPAACVPTTMAAGGAPMADNNEGADGVGNAGNWHCHDSTWLGDRV 239  
Db 181 ESYVDDPOPLGEPPATPAACVPTTMAAGGAPMADNNEGADGVGNAGNWHCHDSTWLGDRV 240  
Qy 240 ITTSTRTWALPTNNHLYKQISSASGAS-NDNHYFGYSTPWPYDFNRRFCHFSPPRDWQ 298  
Db 241 ITTSTRTWALPTNNHLYKQISSASGAS-NDNHYFGYSTPWPYDFNRRFCHFSPPRDWQ 300  
Qy 299 RLNNNGWFPKRLNFKLNIQVKEVTNNEGKTIANNLTSTVQVFTDSEYQLPYVLGSA 358  
Db 301 RLNNNGWFPKRLNFKLNIQVKEVTNNEGKTIANNLTSTVQVFTDSEYQLPYVLGSA 360  
Qy 359 HQGCLPPPPADVFMIPOYGYLTNNGSQAVGRSSFCYCLEYFPSPQMLRTGNNTFESYTFEEV 418  
Db 361 HQGCLPPPPADVFMIPOYGYLTNNGSQAVGRSSFCYCLEYFPSPQMLRTGNNTFESYTFEEV 420  
Qy 419 VPFHSSYAHSSQSLDLRLMNLIDQYLYLNRTOQSGSAQNKDLFLSRGSPAGMSVQPKNW 478  
Db 421 VPFHSSYAHSSQSLDLRLMNLIDQYLYLNRTOQSGSAQNKDLFLSRGSPAGMSVQPKNW 480  
Qy 479 LPGPCYRQORVSKTKTDNNNSNFTWTCASKYNLNGRESIINPGTAMASHKODEDKFPFMS 538  
Db 481 LPGPCYRQORVSKTKTDNNNSNFTWTCASKYNLNGRESIINPGTAMASHKODEDKFPFMS 540  
Qy 539 GVMI FGKESAGASNTALDNLVMTDEBEIKATNPVATERFGTVAVNFPQSSSTDPATGDVHA 598  
Db 541 GVLMFGKQAGKNDVYSSVMTSEEEIKTNPNVATEQYGVVADNLQOQNAAPIVGAVNS 600  
Qy 599 MGALPGMWQDRDYLQGPWAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKNTVPVPANPP 658  
Db 601 QGALPGMWQNRDYLQGPWAKIPHTDGNFHPSPMLMGFGGLKNPPQILIKNTVPVPADP 660  
Qy 659 PAEFSAKTFASFTQYSTQGVSVIEWELOKENSKRNWPEVQYTSNYSKANSANTVDFTDNN 718  
Db 661 PTFESQAKLSNFTQYSTQGVSVIEWELOKENSKRNWPEVQYTSNYSKANSANTVDFAVNT 720  
Qy 719 GLYTEPRPIGTRYLTRPL 736  
Db 721 GTYSEPRPIGTRYLTRNL 738

RESULT 13  
US-10-291-583-79  
; Sequence 79, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; PRIOR FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 79  
; LENGTH: 738  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 44.1  
US-10-291-583-79



Query Match 86.4%; Score 3448; DB 12; Length 738;  
Best Local Similarity 85.0%; Pred. No. 2.2e-291;  
Matches 627; Conservative 46; Mismatches 63; Indels 2; Gaps 2;

Qy 1 MAADGYPDLWLEDNLSGIREWDLKPGAPKPAKQKQDDGRGLVLPYKYLGPFGNGLD 60  
Db 1 MAADGYPDLWLEDNLSGIREWDLKPGAPKPAKQKQDDGRGLVLPYKYLGPFGNGLD 60

Qy 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAERFQERLQEDTSFGNLRGAVFQ 120  
Db 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAERFQERLQEDTSFGNLRGAVFQ 120

Qy 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQ-EPDSSSGIGTKTQGPAPKRLNFGQTGDS 179  
Db 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSPDSTGIGKKGQAPKRLNFGQTGDS 180

Qy 180 ESVPPDQPLGEPAPPAVPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRV 239  
Db 181 ESVPPDQPLGEPAPPAVPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRV 240

Qy 240 ITTSTRTWALPTYNHLYKQIISASTGAS-NDNHYFGYSTPMGYDFNRFCHFPSPDMQ 298  
Db 241 ITTSTRTWALPTYNHLYKQIISNGTSGGSTNDNTYFGYSTPMGYDFNRFCHFPSPDMQ 300

Qy 299 RLNNWGFPRKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFSSEYQLPYVLGSA 358  
Db 301 RLNNWGFPRKRLNFKLFNIQVKEVTNEGKTIANNLTSIQVTFDSEYQLPYVLGSA 360

Qy 359 HOGCLPPPPADVFMIPOYGYLTNNQSGAVGRSSFYCLEYFPFSQMLRTGNFTFSYTFEE 418  
Db 361 HOGCLPPPPADVFMIPOYGYLTNNQSGAVGRSSFYCLEYFPFSQMLRTGNFTFSYQFED 420

Qy 419 VPFHSSYAHQSLSLRLNPLIDQYLYLNRQNSQAKNDLLFSRGSAGMSVQPKNW 478  
Db 421 VPFHSSYAHQSLSLRLNPLIDQYLYLNRQNSQAKNDLLFSRGSAGMSVQPKNW 480

Qy 479 LPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMS 538  
Db 481 LPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMS 540

Qy 539 GVMIFGKESAGASNTALDNVMTDEEIKATNPVATERGTVAVNFQSSSTDPATGDVHA 598  
Db 541 GVMIFGKESAGASNTALDNVMTDEEIKATNPVATERGTVAVNFQSSSTDPATGDVHA 600

Qy 599 MGALPGMWQDRDYLQGPWAKI PHTDGHFPLMGFGFLKPPPPQILLIKNTVPVADP 658  
Db 601 MGALPGMWQDRDYLQGPWAKI PHTDGHFPLMGFGFLKPPPPQILLIKNTVPVADP 660

Qy 659 PAEFSATKFPASFTQYSTQGVSVIEWELQKENSQRWNPVQYTSNYAKSANVDFTVNN 718  
Db 661 PTFESQAKLASFTQYSTQGVSVIEWELQKENSQRWNPVQYTSNYAKSANVDFTVNN 720

Qy 719 GLYTPRPPIGTRYLTRPL 736  
Db 721 GTYSEPRPIGTRYLTRNL 738

RESULT 14  
US-10-291-583-81  
; Sequence 81, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wileon, James M.  
; APPLICANT: Alviria, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; PRIOR FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 81  
; LENGTH: 738  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 44.2  
US-10-291-583-81

Query Match 86.4%; Score 3448; DB 12; Length 738;  
Best Local Similarity 85.0%; Pred. No. 2.2e-291;  
Matches 627; Conservative 46; Mismatches 63; Indels 2; Gaps 2;

Qy 1 MAADGYPDLWLEDNLSGIREWDLKPGAPKPAKQKQDDGRGLVLPYKYLGPFGNGLD 60  
Db 1 MAADGYPDLWLEDNLSGIREWDLKPGAPKPAKQKQDDGRGLVLPYKYLGPFGNGLD 60

Qy 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAERFQERLQEDTSFGNLRGAVFQ 120  
Db 61 KGEVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAERFQERLQEDTSFGNLRGAVFQ 120

Qy 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQ-EPDSSSGIGTKTQGPAPKRLNFGQTGDS 179  
Db 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSPDSTGIGKKGQAPKRLNFGQTGDS 180

Qy 180 ESVPPDQPLGEPAPPAVPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRV 239  
Db 181 ESVPPDQPLGEPAPPAVPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRV 240

Qy 240 ITTSTRTWALPTYNHLYKQIISASTGAS-NDNHYFGYSTPMGYDFNRFCHFPSPDMQ 298  
Db 241 ITTSTRTWALPTYNHLYKQIISNGTSGGSTNDNTYFGYSTPMGYDFNRFCHFPSPDMQ 300

Qy 299 RLNNWGFPRKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFSSEYQLPYVLGSA 358  
Db 301 RLNNWGFPRKRLNFKLFNIQVKEVTNEGKTIANNLTSIQVTFDSEYQLPYVLGSA 360

Qy 359 HOGCLPPPPADVFMIPOYGYLTNNQSGAVGRSSFYCLEYFPFSQMLRTGNFTFSYTFEE 418  
Db 361 HOGCLPPPPADVFMIPOYGYLTNNQSGAVGRSSFYCLEYFPFSQMLRTGNFTFSYQFED 420

Qy 419 VPFHSSYAHQSLSLRLNPLIDQYLYLNRQNSQAKNDLLFSRGSAGMSVQPKNW 478  
Db 421 VPFHSSYAHQSLSLRLNPLIDQYLYLNRQNSQAKNDLLFSRGSAGMSVQPKNW 480

Qy 479 LPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMS 538  
Db 481 LPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMS 540

Qy 539 GVMIFGKESAGASNTALDNVMTDEEIKATNPVATERGTVAVNFQSSSTDPATGDVHA 598  
Db 541 GVMIFGKESAGASNTALDNVMTDEEIKATNPVATERGTVAVNFQSSSTDPATGDVHA 600

Qy 599 MGALPGMWQDRDYLQGPWAKI PHTDGHFPLMGFGFLKPPPPQILLIKNTVPVADP 658  
Db 601 MGALPGMWQDRDYLQGPWAKI PHTDGHFPLMGFGFLKPPPPQILLIKNTVPVADP 660

Qy 659 PAEFSATKFPASFTQYSTQGVSVIEWELQKENSQRWNPVQYTSNYAKSANVDFTVNN 718  
Db 661 PTFESQAKLASFTQYSTQGVSVIEWELQKENSQRWNPVQYTSNYAKSANVDFTVNN 720

Qy 719 GLYTPRPPIGTRYLTRPL 736  
Db 721 GTYSEPRPIGTRYLTRNL 738

RESULT 15  
US-10-291-583-93



; Sequence 93, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; FILE OF INVENTION: Sequences and Isolating No. US20030138772A1e1 Sequences Identif  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 93  
; LENGTH: 738  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 43.12  
US-10-291-583-93

Query Match 86.4%; Score 3445; DB 12; Length 738;  
Best Local Similarity 84.7%; Pred. No. 4e-291;  
Matches 625; Conservative 48; Mismatches 63; Indels 2; Gaps 2;

Qy 1 MAADGYPDLWLEDNLSGIREWMDLPGAPKPKANQKQDDGRGLVLPGYKYLGPFGNLD 60  
Db 1 MAADGYPDLWLEDNLSGIREWMDLPGAPKPKANQKQDDGRGLVLPGYKYLGPFGNLD 60

Qy 61 KGEPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEFQERLOEDTSFGNLRGAVFQ 120  
Db 61 KGEPVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEFQERLOEDTSFGNLRGAVFQ 120

Qy 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQ-EPDSSSG1GKTGOQPAKRLNFGQTGDS 179  
Db 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQSPDSSSTG1GKKHQPAKRLNFGQTGDS 180

Qy 180 ESVDPDPPLGEPPTAPAAVGPPTMASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRV 239  
Db 181 ESVDPDPPLGEPPTAPAAVGPPTMASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRV 240

Qy 240 ITTSTRTWALPTYNNHLYKQISSASTGAS-NDNHYPGYSTPWGYEDFNRPHCHFSPRDMQ 298  
Db 241 ITTSTRTWALPTYNNHLYKQISSASTGAS-NDNHYPGYSTPWGYEDFNRPHCHFSPRDMQ 300

Qy 299 RLINNNGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSA 358  
Db 301 RLINNNGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSA 360

Qy 359 HOGCLPPPPADVFMIPOGYGLTLNNGSOAVGRSSFYCLEYPPSQMLRTGNFTSYTPEE 418  
Db 361 HOGCLPPPPADVFMIPOGYGLTLNNGSOAVGRSSFYCLEYPPSQMLRTGNFTSYTPEE 420

Qy 419 VPFHSSYAHQSLSRLMNPILDOYLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNW 478  
Db 421 VPFHSSYAHQSLSRLMNPILDOYLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNW 480

Qy 479 LPGPCYRQORVSKTTDNNNSFTWTGASKYNLNGRESIINPGTAWASHKDDDEKFFPMS 538  
Db 481 LPGPCYRQORVSKTTDNNNSFTWTGASKYNLNGRESIINPGTAWASHKDDDEKFFPMS 540

Qy 539 GWMIFGKSAGASNTALDNMTDDEETKATNPVATERFGTVAVNFGSSSTDPATGDVHA 598  
Db 541 GWMIFGKSAGASNTALDNMTDDEETKATNPVATERFGTVAVNFGSSSTDPATGDVHA 600

Qy 599 MGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGFGFLKPPQILLIKNTVPANP 658  
Db 599 MGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGFGFLKPPQILLIKNTVPANP 658

Db 601 QGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGFGFLKPPQILLIKNTVPANP 660  
Qy 659 PAEFSATKFAFITQYSTGVSVIEWELQKENSKRWNPEVQYTSNYAKSANVDFTYDNN 718  
Db 661 PTFSSQAKLASFITQYSTGVSVIEWELQKENSKRWNPEIQTSTNYKSTNVDFAVNTE 720  
Qy 719 GLYTEPRPICTRYLTRPL 736  
Db 721 GTYSEPRPICTRYLTRNL 738

Search completed: January 21, 2004, 16:18:23  
Job time : 36.0787 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 15:59:46 ; Search time 44.8925 Seconds  
(without alignments)  
2602.280 Million cell updates/sec

Title: US-09-807-802A-13  
Perfect score: 3989  
Sequence: 1 MAADGYLDWLENDLSEGR.....NNGLYTEPRICTRYLTRPL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues  
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq 13Jun03: \*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: \*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: \*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: \*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: \*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: \*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: \*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: \*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: \*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: \*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: \*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: \*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: \*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: \*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: \*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: \*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: \*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: \*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: \*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: \*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: \*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: \*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: \*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: \*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3989	100.0	736	21	AAV711167
2	3983	99.3	736	22	Adeno-associated v
3	3511	88.0	736	22	AAV6 capsid protei
4	3494	87.6	736	22	AAV3B capsid prote
5	3402.5	85.3	735	22	AAV3A capsid prote
6	3402.5	85.3	735	22	Adeno-associated v
7	3402.5	85.3	735	22	Adeno-associated v
8	3402.5	85.3	735	22	AAV2 capsid protei
9	3376	84.6	734	22	Adeno-associated v

10	3251	81.5	599	21	AAV711168
11	2906	72.9	534	21	Adeno-associated v
12	2759.5	69.2	598	22	Adeno-associated v
13	2759.5	69.2	598	22	AAAG65793
14	2759.5	69.2	598	22	AAV51509
15	2486.5	62.3	734	23	Adeno-associated v
16	2486.5	62.3	734	24	AAV4 VPI capsid pr
17	2481.5	62.2	533	22	Adeno-associated v
18	2481.5	62.2	533	22	AAAG65794
19	2477.5	62.1	533	22	Adeno-associated v
20	2220	55.7	724	21	Adeno-associated v
21	2220	55.7	724	21	AAV58160
22	2195.5	55.0	736	18	Adeno-associated v
23	2168.5	54.4	732	16	Duck parvovirus ca
24	1830.5	45.9	598	19	Barbary duck parvo
25	1830.5	45.9	598	24	AAV4 VP2 coat prot
26	1700.5	42.6	588	21	Adeno-associated v
27	1700.5	42.6	588	21	Adeno-associated v
28	1690.5	42.4	544	24	AAV4 VP3 coat prot
29	1690.5	42.4	544	24	AAV4 VP3 coat prot
30	1676	42.0	534	16	Adeno-associated v
31	1665	41.7	532	21	Barbary duck parvo
32	1665	41.7	532	21	Adeno-associated v
33	1410	35.3	756	21	Adeno-associated v
34	921	23.1	202	23	Capsid protein enc
35	606.5	15.2	781	20	Adeno-associated v
36	605.5	15.2	781	16	Erythrovirus V9 VP
37	599.5	15.0	781	24	Human parvovirus B
38	590.5	14.8	781	24	Human parvovirus B
39	490	12.3	554	16	Human parvovirus V
40	490	12.3	554	20	Human parvovirus V
41	486	12.2	554	24	Erythrovirus V9 VP
42	486	12.2	554	24	Adeno-associated v
43	476.5	11.9	543	12	Human parvovirus B
44	425.5	10.7	472	5	Parvo virus B19 VP
45	417.5	10.5	620	22	Sequence encoded b
					Feline parvovirus

ALIGNMENTS

RESULT 1  
AAV711167  
ID AAV711167 standard; Protein; 736 AA.  
XX  
AC AAV711167;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 capsid protein VPI.  
XX  
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VPI.  
XX  
OS Adeno associated virus serotype 1.  
XX  
PN WO200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
DR WPI; 2000-376571/32.  
XX  
DR N-PSDB; AAD00772, AAD00777.  
XX  
PT Novel adeno-associated virus serotype 1 polynucleotide useful for

PT preparation of medicament for delivery of a transgene to a host -

PS Claim 7; Page 87-90; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VPI which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX SQ Sequence 736 AA;

Query Match 100.0%; Score 3989; DB 21; Length 736;  
Best Local Similarity 100.0%; Pred. No. 1e-311;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAADGYPDLWEDNLSEGIREWMDLPGAPKPKANQKQDDGRGLVLPYKYLGPFGNGLD 60  
DB 1 MAADGYPDLWEDNLSEGIREWMDLPGAPKPKANQKQDDGRGLVLPYKYLGPFGNGLD 60

QY 61 KGPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRVAVQ 120  
DB 61 KGPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRVAVQ 120

QY 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSDSSGGIGKTGQOPAKKRLNFGTGDSE 180  
DB 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSDSSGGIGKTGQOPAKKRLNFGTGDSE 180

QY 181 SVPDPQPLGEPPTAAVGTFTMASGGGAPMADNNEGADVGNASGNHCHDSTWLGDRVI 240  
DB 181 SVPDPQPLGEPPTAAVGTFTMASGGGAPMADNNEGADVGNASGNHCHDSTWLGDRVI 240

QY 241 TTSTRTWALPTNNHLYKQISSASTGASNDNHFGYSTPWGTFDNRHCHSPSDWQRL 300  
DB 241 TTSTRTWALPTNNHLYKQISSASTGASNDNHFGYSTPWGTFDNRHCHSPSDWQRL 300

QY 301 INNMGFRPKRLNFKLFNIQVKEVTNDGVTIANNLTSTVQVFSDESQYLPYVLSAHQ 360  
DB 301 INNMGFRPKRLNFKLFNIQVKEVTNDGVTIANNLTSTVQVFSDESQYLPYVLSAHQ 360

QY 361 GCLPPPPADVPMI PQGYLTLLNGSAQVGRSSFYCLEYPPSQMLRTGNNFTSYTFEEVP 420  
DB 361 GCLPPPPADVPMI PQGYLTLLNGSAQVGRSSFYCLEYPPSQMLRTGNNFTSYTFEEVP 420

QY 421 FHSVAHSOSLRLMNLIDQYLYLNRNTQNSGSAQNKDILFSGSPAGMSVQPKNMLP 480  
DB 421 FHSVAHSOSLRLMNLIDQYLYLNRNTQNSGSAQNKDILFSGSPAGMSVQPKNMLP 480

QY 481 GPCYRQORVSKTKTNNNSNFTWTGASKYNLNGRESII NFGTAMASHKDDKFFPMSCV 540  
DB 481 GPCYRQORVSKTKTNNNSNFTWTGASKYNLNGRESII NFGTAMASHKDDKFFPMSCV 540

QY 541 MIFGKESAGASNTALDNVMTDEEETKATNPVATERFGTVAVNFOSSSTDPATGDVHAMG 600  
DB 541 MIFGKESAGASNTALDNVMTDEEETKATNPVATERFGTVAVNFOSSSTDPATGDVHAMG 600

QY 601 ALPGMWQDRDYYLQGPVIAKIPHTDGHFHPSPLMGFGFKNPPPOILLIKNTPVPANPPA 660  
DB 601 ALPGMWQDRDYYLQGPVIAKIPHTDGHFHPSPLMGFGFKNPPPOILLIKNTPVPANPPA 660

QY 661 EFSATKFASTFYQYSTGVSVIEWELQENSKRWNPVEQVTSNYAKSANVFTVDNNGL 720  
DB 661 EFSATKFASTFYQYSTGVSVIEWELQENSKRWNPVEQVTSNYAKSANVFTVDNNGL 720

QY 721 YTEPRPIGTRYLTRPL 736  
DB 721 YTEPRPIGTRYLTRPL 736

# RESULT 2

AAB59847

ID AAB59847 standard; Protein; 736 AA.

XX AC AAB59847;

XX DT 28-MAR-2001 (first entry)

XX DE AAV6 capsid protein VPI.

XX KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;

XX KW atherosclerosis; sickle cell anaemia; thalassaemia;

XX KW blood clotting disorder; diabetes; capsid protein VPI.

XX OS Adeno associated virus.

XX PN US6156303-A.

XX PD 05-DEC-2000.

XX PF 11-JUN-1997; 97US-0873168.

XX PR 11-JUN-1997; 97US-0873168.

XX PA (UNITW ) UNIV WASHINGTON.

XX PI Russell DW, Rutledge BA;

XX DR WPI; 2001-060164/07.

XX PT Adeno-associated virus serotype 6 and viral vector derived from it for

XX PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency

XX PS syndrome, sickle cell anemia, thalassaemia and diabetes -

XX PS Disclosure; Fig 2; 50pp; English.

XX CC The present invention relates to adeno-associated virus serotypes. The

XX CC present sequence is capsid protein VPI of one such serotype (AAV6). AAV6

XX CC can be used to construct AAV viral vectors for use in gene therapy for a

XX CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,

XX CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.

XX CC The AAV viral vectors have increased transduction efficiency of a

XX CC particular host cell as the AAV virion containing the AAV vector genome

XX CC can be modified to express a capsid protein of an AAV serotype that

XX CC transduces the selected host cell.

XX SQ Sequence 736 AA;

Query Match 99.3%; Score 3963; DB 22; Length 736;

Best Local Similarity 99.2%; Pred. No. 1.2e-309;

Matches 730; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAADGYPDLWEDNLSEGIREWMDLPGAPKPKANQKQDDGRGLVLPYKYLGPFGNGLD 60

DB 1 MAADGYPDLWEDNLSEGIREWMDLPGAPKPKANQKQDDGRGLVLPYKYLGPFGNGLD 60

QY 61 KGPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRVAVQ 120

DB 61 KGPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLRVAVQ 120

QY 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSDSSGGIGKTGQOPAKKRLNFGTGDSE 180

DB 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSDSSGGIGKTGQOPAKKRLNFGTGDSE 180

QY 181 SVPDPQPLGEPPTAAVGTFTMASGGGAPMADNNEGADVGNASGNHCHDSTWLGDRVI 240

DB 181 SVPDPQPLGEPPTAAVGTFTMASGGGAPMADNNEGADVGNASGNHCHDSTWLGDRVI 240

CC The present invention relates to adeno-associated virus serotypes. The present sequence is capsid protein VP1 of one such serotype (AAV3B). AAV3B can be used to construct AAV viral vectors for use in gene therapy for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis, sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes. The AAV viral vectors have increased transduction efficiency of a particular host cell as the AAV virion containing the AAV vector genome can be modified to express a capsid protein of an AAV serotype that transduces the selected host cell.

CC  
XX  
SQ

Query Match  
Best Local Similarity 88.0%; Score 3511; DB 22; Length 736;  
Matches 640; Conservative 38; Mismatches 57; Indels 2; Gaps 2;

Qy 1 MAADGVLPMLENDNLSEGIKREWDLKPGAPKPKANQKQDGRGLVLPYKYLPGPNGLD 60  
Db 1 MAADGVLPMLENDNLSEGIKREWDLKPGAPKPKANQKQDGRGLVLPYKYLPGPNGLD 60

Qy 61 KGEVNVNADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLQEDTSFGNGLGRAVFQ 120  
Db 61 KGEVNVNADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLQEDTSFGNGLGRAVFQ 120

Qy 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSPDSSSGIGTKQQQPAKRLNFGQTDSE 180  
Db 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSPDSSSGIGTKQQQPAKRLNFGQTDSE 180

Qy 181 SYPDQPLGEPAATPAVQPTTMSGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
Db 181 SYPDQPLGEPAATPAVQPTTMSGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240

Qy 241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPWGYFDNRFCHFSPRDQRL 300  
Db 241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPWGYFDNRFCHFSPRDQRL 300

Qy 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTIIANNLTSTQVFSSEYQLPVYLSAHQ 360  
Db 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTIIANNLTSTQVFSSEYQLPVYLSAHQ 360

Qy 361 GCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPFQMLRTGNNTFSYTFEVP 420  
Db 361 GCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPFQMLRTGNNTFSYTFEVP 420

Qy 421 FHSSYAHQSQDLRLNPLIDQYLYLNRTQSGSAQNKDLFSRSGSPAGMSVQPKNWL 480  
Db 421 FHSSYAHQSQDLRLNPLIDQYLYLNRTQSGSAQNKDLFSRSGSPAGMSVQPKNWL 480

Qy 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAWASHKDDKDFPMSGV 540  
Db 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAWASHKDDKDFPMSGV 540

Qy 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATEREGTVAVNFQSSSTDPATGDVHAMG 600  
Db 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATEREGTVAVNFQSSSTDPATGDVHAMG 600

Qy 601 ALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGSGFLGNKPPQILIKNTVPANPPA 660  
Db 601 ALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGSGFLGNKPPQILIKNTVPANPPA 660

Qy 661 EFSATKFPASFIQYSTQGVSEIWELOKENSXRNWPEVQYTSNYAKSANVDFTVDNNG 720  
Db 661 EFSATKFPASFIQYSTQGVSEIWELOKENSXRNWPEVQYTSNYAKSANVDFTVDNNG 720

Qy 721 YTEPRPIGTRYLTRPL 736  
Db 721 YTEPRPIGTRYLTRPL 736

RESULT 3  
AAB59846  
ID AAB59846 standard; Protein; 736 AA.

XX  
AC AAB59846;  
XX  
XX 28-MAR-2001 (first entry)  
DT  
DE AAV3B capsid protein VP1.  
XX  
XX AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
XX atherosclerosis; sickle cell anaemia; thalassaemia;  
XX blood clotting disorder; diabetes; capsid protein VP1.  
XX  
XX Adeno associated virus.  
XX  
XX US6156303-A.  
XX  
XX 05-DEC-2000.  
XX  
XX 11-JUN-1997; 97US-0873168.  
XX  
XX 11-JUN-1997; 97US-0873168.  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Russell DW, Rutledge EA;  
XX WPI; 2001-060164/07.  
XX  
XX Adeno-associated virus serotype 6 and viral vector derived from it for  
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
XX syndrome, sickle cell anemia, thalassaemia and diabetes  
XX  
XX Disclosure; Fig 2; 50pp; English.

CC The present invention relates to adeno-associated virus serotypes. The present sequence is capsid protein VP1 of one such serotype (AAV3B). AAV3B can be used to construct AAV viral vectors for use in gene therapy for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis, sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes. The AAV viral vectors have increased transduction efficiency of a particular host cell as the AAV virion containing the AAV vector genome can be modified to express a capsid protein of an AAV serotype that transduces the selected host cell.

CC  
XX  
SQ

Query Match  
Best Local Similarity 86.8%; Pred. No. 2.8e-273;  
Matches 640; Conservative 38; Mismatches 57; Indels 2; Gaps 2;

Qy 1 MAADGVLPMLENDNLSEGIKREWDLKPGAPKPKANQKQDGRGLVLPYKYLPGPNGLD 60  
Db 1 MAADGVLPMLENDNLSEGIKREWDLKPGAPKPKANQKQDGRGLVLPYKYLPGPNGLD 60

Qy 61 KGEVNVNADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLQEDTSFGNGLGRAVFQ 120  
Db 61 KGEVNVNADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLQEDTSFGNGLGRAVFQ 120

Qy 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSPDSSSGIGTKQQQPAKRLNFGQTDSE 180  
Db 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQSPDSSSGIGTKQQQPAKRLNFGQTDSE 180

Qy 181 SYPDQPLGEPAATPAVQPTTMSGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
Db 181 SYPDQPLGEPAATPAVQPTTMSGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240

Qy 241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPWGYFDNRFCHFSPRDQRL 300  
Db 241 TTSTRTWALPTYNHLYKQISSASTGASNDNHFGYSTPWGYFDNRFCHFSPRDQRL 300

Qy 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTIIANNLTSTQVFSSEYQLPVYLSAHQ 360  
Db 301 INNNWGFPRKRLNFKLFNIQVKEVTNDGVTIIANNLTSTQVFSSEYQLPVYLSAHQ 360

Qy 361 GCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPFQMLRTGNNTFSYTFEVP 420  
Db 361 GCLPPFPADVFMIPOGYLTLNNGSQAVGRSSFCLEYFPFQMLRTGNNTFSYTFEVP 420

Qy 421 FHSSYAHQSQDLRLNPLIDQYLYLNRTQSGTTSQTTSQTTSQTTSQTTSQTTSQTTSQTTS 479  
Db 421 FHSSYAHQSQDLRLNPLIDQYLYLNRTQSGTTSQTTSQTTSQTTSQTTSQTTSQTTSQTTS 479

Qy 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAWASHKDDKDFPMSGV 539  
Db 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAWASHKDDKDFPMSGV 539

Qy 540 VMI FKGESAGASNTALDNVMTDDEEIKATNPVATEREGTVAVNFQSSSTDPATGDVHAM 599  
Db 540 VMI FKGESAGASNTALDNVMTDDEEIKATNPVATEREGTVAVNFQSSSTDPATGDVHAM 599

Qy 600 GALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGSGFLGNKPPQILIKNTVPANPP 659  
Db 600 GALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGSGFLGNKPPQILIKNTVPANPP 659

Qy 660 AEFSAKFPASFIQYSTQGVSEIWELOKENSXRNWPEVQYTSNYAKSANVDFTVDNNG 719  
Db 660 AEFSAKFPASFIQYSTQGVSEIWELOKENSXRNWPEVQYTSNYAKSANVDFTVDNNG 719

Qy 720 LYTEPRPIGTRYLTRPL 736  
Db 720 LYTEPRPIGTRYLTRPL 736

RESULT 4  
AAB59845  
ID AAB59845 standard; Protein; 736 AA.

XX

AAB59845;  
28-MAR-2001 (first entry)  
AAV3A capsid protein VP1.  
AAV3A; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
atherosclerosis; sickle cell anaemia; thalassaemia;  
blood clotting disorder; diabetes; capsid protein VP1.  
Adeno associated virus.  
US6156303-A.  
05-DEC-2000.  
11-JUN-1997; 97US-0873168.  
11-JUN-1997; 97US-0873168.  
(UNIW ) UNIV WASHINGTON.  
Russell DW, Rutledge EA;  
WPI; 2001-060164/07.  
Adeno-associated virus serotype 6 and viral vector derived from it for  
gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
syndrome, sickle cell anemia, thalassaemia and diabetes -  
Disclosure; Fig 2; 50pp; English.  
The present invention relates to adeno-associated virus serotypes. The  
present sequence is capsid protein VP1 of one such serotype (AAV3A).  
AAV3A can be used to construct AAV viral vectors for use in gene therapy  
for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
The AAV viral vectors have increased transduction efficiency of a  
particular host cell as the AAV virion containing the AAV vector genome  
can be modified to express a capsid protein of an AAV serotype that  
transduces the selected host cell.  
Sequence 736 AA;  
Query Match 87.6%; Score 3494; DB 22; Length 736;  
Best Local Similarity 86.4%; Pred. No. 6.6e-272;  
Matches 637; Conservative 39; Mismatches 59; Indels 2; Gaps 2;  
QY 1 MAADGYLPDWLENDLSEGREWDLKPGAPKPKANQKQDDGRLVLPKYKYLGPNGLD 60  
DB 1 MAADGYLPDWLENDLSEGREWDLKPGAPKPKANQKQDDGRLVLPKYKYLGPNGLD 60  
QY 61 KGEPVNAADAALAHDKAYDQQLKAGDNPYLRYNHADAFFQERLQEDTSFGNGLGRAVQ 120  
DB 61 KGEPVNEADAALAHDKAYDQQLKAGDNPYLRYNHADAFFQERLQEDTSFGNGLGRAVQ 120  
QY 121 AKKRVLEPLGLVEEGAKTAPGKRVPVEQSPQEPDSSSGIGTKTQQPAKRLNFGQTGDSE 180  
DB 121 AKKRVLEPLGLVEEGAKTAPGKRVPVEQSPQEPDSSSGIGTKTQQPAKRLNFGQTGDSE 180  
QY 181 SVDPDQPLGEPPATPAAGVPTTASGGGAPMADNNEGADGVGNAGNCHDSTWLGDRVI 240  
DB 181 SVDPDQPLGEPPATPAAGVPTTASGGGAPMADNNEGADGVGNAGNCHDSTWLGDRVI 240  
QY 241 TTSTRTWALPTYNHLYKQISSASTCASNDNHFGYSTPMGYDFNRFCHFSPRDWORL 300  
DB 241 TTSTRTWALPTYNHLYKQISSASTCASNDNHFGYSTPMGYDFNRFCHFSPRDWORL 299  
QY 301 INNWGFRPKRLNFKLFIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQ 360  
DB 300 INNWGFRPKRLNFKLFIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQ 359  
QY 361 GCLPPFPADVFMIPQYGYLTLLNGSQAVGRSSFYCLEYFPQMLRTGNFQSYTFEDVP 420

DB 360 GCLPPFPADVFMIPQYGYLTLLNGSQAVGRSSFYCLEYFPQMLRTGNFQSYTFEDVP 419  
QY 421 FHSSVAHSQSLDRMLNPLIDQLYLNRTO-NQSGSAQNKDLLFSRGSPAGMSVQPKNWL 479  
DB 420 FHSSVAHSQSLDRMLNPLIDQLYLNRTOGTTTSGTTNQSRLLSQAGPQMSLOARNWL 479  
QY 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLANGRESIINFGTAWASHKDDKPFPMG 539  
DB 480 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLANGRESIINFGTAWASHKDDKPFPMG 539  
QY 540 VMIFKESAGASNTALONVMTDDEEIKATNPVATERGTAVNFQSSSTDPATGDVHAM 599  
DB 540 NLIKFGKGTASNAELDNVMTDDEEIRTTNPVATEQYGTVANLQSSNTATTCTGVNHQ 599  
QY 600 GALPGVMWQDRDVLQGPPIWAKIPHTDGHFHPSPLMGGFGLKNNPPQILIKNTVPANPP 659  
DB 600 GALPGVMWQDRDVLQGPPIWAKIPHTDGHFHPSPLMGGFGLKNNPPQILIKNTVPANPP 659  
QY 660 AEFSAKFPASFTITQYSTGQSVSEIWELOKNSKRNPEVQYTSNYAKSANVDFTVDNNG 719  
DB 660 TTFSPAKFPASFTITQYSTGQSVSEIWELOKNSKRNPEIQTSTNYKSNVNVDFTVDTNG 719  
QY 720 LYTPRPPIGTRYLTPPL 736  
DB 720 VYSPRPPIGTRYLTPPL 736  
RESULT 5  
AAG65792 standard; Protein; 735 AA.  
XX AAG65792;  
XX 11-FEB-2002 (first entry)  
XX Adeno-associated virus 2 (AAV-2) major coat protein VP1.  
XX Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
XX inverted terminal repeat; neurotropic; neuroprotective; antianemic; ITR;  
XX antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
XX major coat protein; AAV-2; VP1.  
XX Adeno-associated virus 2.  
XX WO200168888-A2.  
XX 20-SEP-2001.  
XX 13-MAR-2001; 2001WO-US07927.  
XX 14-MAR-2000; 2000US-189110P.  
XX (NEUR-) NEUROLOGIX INC.  
XX Xiao W, During MJ;  
XX WPI; 2001-596912/67.  
XX N-PSDB; AAI66974.  
XX Recombinant viral vector useful in improving gene therapy in a subject,  
XX and for increasing efficiency of entry into a cell, comprises a  
XX chimeric capsid having one non-native amino acid sequence and a desired  
XX transgene -  
XX Disclosure; Page 50; 53pp; English.  
XX The invention provides a recombinant viral vector (RVV) comprising a  
XX chimeric capsid (I) having at least one non-native amino acid sequence,  
XX derived from a capsid protein domain of parvovirus (II), a virus (III),  
XX or their combination, and a transgene flanked 5' and 3' by inverted  
XX terminal repeat (ITR) sequences, derived from (II), (III), or their  
XX combination. The RVV is useful for improving gene therapy in a subject

with a disorder, and for increasing the efficiency of entry into a cell, which involves producing (I) encapsulating a viral vector, and contacting a cell with RVW having (I) such that (I) binds to an attachment site on the cell surface and permits the vector to enter the cell efficiently. A pharmaceutical composition comprising RVW with (I) containing a transgene sequence associated with a disease or a disorder such that expression of the transgene would result in amelioration of the disease or disorder such as inherited neurological and metabolic diseases e.g. lysosomal storage disease, Leish-Nyhan syndrome, amyloid polyneuropathy, Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic fibrosis, diabetes, diseases associated with hormone deficiencies, retinoblastoma and various types of neoplastic cells which include, tumours especially central nervous system tumours, neoplasms, carcinomas, sarcomas, leukemias and lymphoma. The present sequence represents the adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences are used in the construction of a chimeric vector.

Seq	Sequence	735 AA;
	Query Match 85.3%; Score 3402.5; DB 22; Length 735; Best Local Similarity 83.3%; Pred. No. 1.5e-264; Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;	
Qy	1 MAADGYLPDWLEEDNLSGIRIEWDLKPGAPKPKANQKQDDGRGLVLPYKYKLGPFNGLD 60	
Db	1 MAADGYLPDWLEEDLTSIGIRIOWMKLKPGPPPKPAERHKDSDRGLVLPYKYKLGPFNGLD 60	
Qy	61 KGEVNAADAAALBHDKAYDOOLKAGDNPYLYRNHADAEOERLQEDTSGCNGLGRAVFQ 120	
Db	61 KGEVNEADAALBHDKAYDQLDSGDNPYLYRNHADAEOERLQEDTSGCNGLGRAVFQ 120	
Qy	121 AKKRVLPGLVEEGAKTAPGKGRPVQSPQEPDSSGGIGXTGQOPAKRLNFGQTGDSE 180	
Db	121 AKKRVLPGLVEBPVKTAPGKGRPVVHS PVEPDSSSGTGKAGQOPARKLNFGQTGDAD 180	
Qy	181 SVDPQPLGEBPATPAAGVPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVI 240	
Db	181 SVDPQPLGQBPAPAGSLGNTMTAGAPMADNNEGADGVGNSGNGWHCDSTWGMGRVI 240	
Qy	241 TTSRTTVALPTYNHLYKQISSASTGASDNHYPGYSTPMWGYDFNRFCHFSPRDWORL 300	
Db	241 TTSRTTVALPTYNHLYKQISSQS- GASDNHYPGYSTPMWGYDFNRFCHFSPRDWORL 299	
Qy	301 INNMGFRPKRLNPKLFNIQVKEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360	
Db	300 INNMGFRPKRLNPKLFNIQVKEVTQNDGTTIANNLTTSTVQVFTDSEYQLPYVLGSAHQ 359	
Qy	361 GCLPFPFADVPMIPOYGYLTLLNQSQAVGRSSFYCLEYFPPSOMLRTGNNTFSTFEVEVP 420	
Db	360 GCLPFPFADVPMVPOYGYLTLLNQSQAVGRSSFYCLEYFPPSOMLRTGNNTFSTFEVDVP 419	
Qy	421 FHSSYAHQSQSLDRLMNPLIDQYLYLARTQNGSAQNKKLLPSRGS PAGMSVPQKNWLP 480	
Db	420 FHSSYAHQSQSLDRLMNPLIDQYLYLSRTNTPSGTTTQSRLLQFSQAGASDIRDQSRNWL 479	
Qy	481 GPCYRQORVSKTKTDNNNSFTWTGASKYNLNGRESIINFGTAMASHKDDDEKFPFMSGV 540	
Db	480 GPCYRQORVSKTSDANNNSSEWTGATKYHLNGRDSLVPNPGPAMASHKDDDEKFPMSGV 539	
Qy	541 MIFKESAGASNTALDNVMIITDEEIEKATNPVATERTGTVAVNFQSSSTDPATGDVHAMG 600	
Db	540 LIFKQSGSEKTNVDIEKVMITDEEIRTTNPVATEQYGSVSTNLQGRNQAAATADVNTQG 599	
Qy	601 ALPGMWQDRDYLQGPITWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTPTVPANPPA 660	
Db	600 VLPGMWQDRDYLQGPITWAKIPHTDGHFHPSPLMGGFGLKHPPPQILIKNTPTVPANPST 659	
Qy	661 EFSATKASFTQYSTQGVSEIWEELQKENS KRNPEVQYTSNYAKSANTDFTVDNGL 720	
Db	660 TFSAAKASFTQYSTQGVSEIWEELQKENS KRNPEIQYTSNYNKS VNVDFTVDNGV 719	
Qy	721 YTEPRPIGTRYLTRPL 736	

```

Db      720 YSBRPIGTYLTRL 735
      |:|||||
RESULT 6
AAMS1508
ID    AAMS1508 standard; protein; 735 AA.
XX
XX    AAMS1508;
XX
XX    02-JAN-2002 (first entry)
XX
XX    Adeno-associated virus VP1 capsid protein.
XX
XX    Adeno-associated virus; AAV; VP1; capsid; virus-like particle;
KW    nuclear localisation signal; VP3.
XX
XX    Adeno associated virus.
XX
XX    JP2001169777-A.
XX
XX    26-JUN-2001.
XX
XX    30-JUL-1999; 99JP-0249140.
XX
XX    30-JUL-1999; 99JP-0249140.
XX
XX    (HAND/) HANDA H.
XX
XX    WPI; 2001-599854/68.
XX
XX    New virus-like particles from VP3 capsid protein of adeno-associated
PT    virus, comprise a peptide containing a nucleus-shifting signal
PT    connected to its N-terminal -
XX
XX    Disclosure; Page 10-13; 33pp; Japanese.
XX
XX    The present sequence is provided in a specification relating to a
CC    virus-like particle-forming protein, and to a peptide containing a
CC    nuclear-shifting signal at its N-terminus that can form virus-like
CC    particles by shifting to the nucleus of the animal cell in which it
CC    is expressed. the method is used for forming virus-like particles
CC    from the VP3 capsid protein of adeno-associated virus (AAV).
XX
XX    Sequence 735 AA;

Query Match      85.1%; Score 3402.5; DB 22; Length 735;
Best Local Similarity 83.3%; Pred. No. 1.5e-264;
Matches 61; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

Qy 1 MAADGYLPDWLEDLSEGIREMMDLKGAPKPKANOOKDDGRGLVLPGYKYLGPFGNGLD 60
Db 1 MAADGYLPDWLEDTLSEGINQWMLKFGPPPPKPAERHKDDSRGLVLPGYKYLGPFGNGLD 60
Qy 61 KGEFVNADAAALEHDKAYDQOLKAGDNPYLRYNHADAEPQERLQEDTSFGGNLGRAVFQ 120
Db 61 KGEFVNADAAALEHDKAYDQOLDGDNPYLYKYNHADAEPQERLKEBTSFGGNLGRAVFQ 120
Qy 121 AKKRVLEPLGLVEEGAKTAPKKRPPVEQSPQESDSSSGIGKTGQOPAKKRLNFGQTGDSE 180
Db 121 AKKRVLEPLGLVEEPVKYAPKKRPVHSPVEPDSSSGTGKAGQOPAKKRLNFGQTGDAD 180
Qy 181 SVDPDQPLGPPATPAAGVPTTWMASGGAPMADNNEGADGVGNASGNHCDSTWLGRDVI 240
Db 181 SVDPDQPLGPPAAPSLGTTNTWATGSGAPMADNNEGADGVGNSSGNHCDSTWMGDRVI 240
Qy 241 TTSTRTWALPTYNNHLYKQISSASTCASNDNHVFGYSTPMGYDFDNRPHCHESPRDQRL 300
Db 241 TTSTRTWALPTYNNHLYKQISSOS - GASNDNHVFGYSTPMGYDFDNRPHCHESPRDQRL 299
Qy 301 INNWFGRPKRLNFKLFNIQKEVTNTDGVTTIANNLTSVQVFSSEYQLPYVLGSAHQ 360
Db 300 INNWFGRPKRLNFKLFNIQKEVTNDGTTTIANNLTSVQVFTSEYQIPYVLSAHO 359

```

QY 361 GCLPPFPADVFMIPQYGYLTLLNGSQAVGRSSFYCLEYFPSQMLRTGNNFTSYTFEEVP 420  
 DB 360 GCLPPFPADVFMIPQYGYLTLLNGSQAVGRSSFYCLEYFPSQMLRTGNNFTSYTFEDVP 419  
 QY 421 FHSSYAHQSLSLDRMLNPLIDQYLYLNRQNSQAQKOLLPSRGSPAGMSVQPKWLP 480  
 DB 420 FHSSYAHQSLSLDRMLNPLIDQYLYLNRQNSQAQKOLLPSRGSPAGMSVQPKWLP 479  
 QY 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYLNNGRESIINPGTAMASHKDDKDEKFFPMGSV 540  
 DB 480 GPCYRQQRVSKTSADNNNSSEYVSWTATKYHLNGRDSLNVNPGPAMASHKDDKDEKFFPMGSV 539  
 QY 541 MIFGKESAGASNTALDNNMTDEBEIKATNPVATERFCTVAVNPQSSSTDPATGDVHAMG 600  
 DB 540 LIFGKQSGSEKTNVDIEKVMITDEBEIRTNVPATEQYGSVSTNLQRGNRQAATADVNTQG 599  
 QY 601 ALPGMWQDRDVIYLGQPIWAKI PHTDGHFHPSPMLGGFGLKNPPOQLIKNTVPANPPA 660  
 DB 600 VLPQMWQDRDVIYLGQPIWAKI PHTDGHFHPSPMLGGFGLKNPPOQLIKNTVPANPST 659  
 QY 661 EFSATKPFASFITQYSTQGVSVIEIWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNG 720  
 DB 660 TFSAAKPFASFITQYSTQGVSVIEIWELOKENSKRWNPEIQYTSNYKSVNVDFVDTNGV 719  
 QY 721 YTEPRPIGTRYLTRPL 736  
 DB 720 YSEPRPIGTRYLTRNL 735

## RESULT 7

AA059844  
 ID AA059844 standard; Protein; 735 AA.

AC AA059844;

DT 28-MAR-2001 (first entry)

DE AAV2 capsid protein VP1.

KW AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
 atherosclerosis; sickle cell anaemia; thalassaemia;  
 blood clotting disorder; diabetes; capsid protein VP1.

OS Adeno associated virus.

PN US6156303-A.

XX 05-DEC-2000.

PF 11-JUN-1997; 97US-0873168.

PR 11-JUN-1997; 97US-0873168.

XX (UNIW ) UNIV WASHINGTON.

XX Russell DW, Rutledge EA;

PI WPI; 2001-060164/07.

PT Adeno-associated virus serotype 6 and viral vector derived from it for  
 gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
 syndrome, sickle cell anaemia, thalassaemia and diabetes -

PS Claim 7; Fig 2; 50pp; English.

XX The present invention relates to adeno-associated virus serotypes. The  
 CC present sequence is capsid protein VP1 of one such serotype (AAV2). AAV2  
 CC can be used to construct AAV viral vectors for use in gene therapy for a  
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
 CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
 CC The AAV viral vectors have increased transduction efficiency of a  
 CC particular host cell as the AAV virion containing the AAV vector genome

CC can be modified to express a capsid protein of an AAV serotype that  
 transduces the selected host cell.

XX Sequence 735 AA;

Query Match 85.3%; Score 3402.5; DB 22; Length 735;  
 Best Local Similarity 83.3%; Pred. No. 1.5e-264;  
 Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPDLENLSEGIEMWDLKPGAPKPKANQKQDDGRGLVLPGYKYLGPNGLD 60

DB 1 MAADGYLPDLENLSEGIEMWDLKPGAPKPKANQKQDDGRGLVLPGYKYLGPNGLD 60

QY 61 KGEFVNAADAALAEHDKAYDQOLKAGDNPVLRVNHADASFOELQEDTSPGGNLGRAVFO 120

DB 61 KGEFVNAADAALAEHDKAYDQOLKAGDNPVLRVNHADASFOELQEDTSPGGNLGRAVFO 120

QY 121 AKRVLLEPLGLVEGAKTAPKPRVQEQSPQBPDSGSGIGKTCQQAQKELNFGQTGDS 180

DB 121 AKRVLLEPLGLVEGAKTAPKPRVQEQSPQBPDSGSGIGKTCQQAQKELNFGQTGDS 180

QY 181 SVDPDQPLGEPPTAPAAVGTPTWASGGGAPMADNNEGADGVNAGNWHCDSTWLDGRVI 240

DB 181 SVDPDQPLGEPPTAPAAVGTPTWASGGGAPMADNNEGADGVNAGNWHCDSTWLDGRVI 240

QY 241 TTSTRTWALFTYNNHLYKQISSASTGASNDNHYFGYSTPMGYDFNRFCHFSFPRDQRL 300

DB 241 TTSTRTWALFTYNNHLYKQISSASTGASNDNHYFGYSTPMGYDFNRFCHFSFPRDQRL 300

QY 301 INNNWGPRLKPLNFKNFNIQVKEVTNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHQ 360

DB 300 INNNWGPRLKPLNFKNFNIQVKEVTNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHQ 360

QY 361 GCLPPFPADVFMIPQYGYLTLLNGSQAVGRSSFYCLEYFPSQMLRTGNNFTSYTFEEVP 420

DB 360 GCLPPFPADVFMIPQYGYLTLLNGSQAVGRSSFYCLEYFPSQMLRTGNNFTSYTFEEVP 420

QY 421 FHSSYAHQSLSLDRMLNPLIDQYLYLNRQNSQAQKOLLPSRGSPAGMSVQPKWLP 480

DB 420 FHSSYAHQSLSLDRMLNPLIDQYLYLNRQNSQAQKOLLPSRGSPAGMSVQPKWLP 480

QY 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYLNNGRESIINPGTAMASHKDDKDEKFFPMGSV 540

DB 480 GPCYRQQRVSKTSADNNNSSEYVSWTATKYHLNGRDSLNVNPGPAMASHKDDKDEKFFPMGSV 539

QY 541 MIFGKESAGASNTALDNNMTDEBEIKATNPVATERFCTVAVNPQSSSTDPATGDVHAMG 600

DB 540 LIFGKQSGSEKTNVDIEKVMITDEBEIRTNVPATEQYGSVSTNLQRGNRQAATADVNTQG 599

QY 601 ALPGMWQDRDVIYLGQPIWAKI PHTDGHFHPSPMLGGFGLKNPPOQLIKNTVPANPPA 660

DB 600 VLPQMWQDRDVIYLGQPIWAKI PHTDGHFHPSPMLGGFGLKNPPOQLIKNTVPANPST 659

QY 661 EFSATKPFASFITQYSTQGVSVIEIWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNG 720

DB 660 TFSAAKPFASFITQYSTQGVSVIEIWELOKENSKRWNPEIQYTSNYKSVNVDFVDTNGV 719

QY 721 YTEPRPIGTRYLTRPL 736

DB 720 YSEPRPIGTRYLTRNL 735

## RESULT 8

AA098974  
 ID AA098974 standard; Protein; 735 AA.

XX AA098974;

AC AA098974;

DT 05-NOV-2002 (first entry)

XX Adeno-associated virus 2 (AAV2) vector, VP1 capsid protein.

XX Adeno-associated virus 2 vector; AAV2; cancer; VP1 capsid;





CC oligopeptides may further be used to bind to host antibody to provide a  
CC transient tolerant or non-responsive state.

SQ Sequence 734 AA;  
Query Match 84.6%; Score 3376; DB 22; Length 734;  
Best Local Similarity 83.0%; Pred. No. 2e-262;  
Matches 611; Conservative 51; Mismatches 72; Indels 2; Gaps 2;  
QY 1 MAADGYLPDWLENDLSEGIREWMDLPGAPKPKANQKQDDGRLGLVLPYKYLGFNFGLD 60  
DB 1 MAADGYLPDWLENDLSEGIROWKWKLPKPPPPKPAERHKDDSRGLVLPYKYLGFNFGLD 60  
QY 61 KGEVNAADAALAEHDKAYDQOLKAGDNPYLYRNHADAERFQERLOEDTSFGCNLGRAVFQ 120  
DB 61 KGEVNEADAALAEHDKAYDRQLDSDGNPYLYRNHADAERFQERLKEDTSFGCNLGRAVFQ 120  
QY 121 AKKRVLEPLGLVEGAKTAPGKRPVEQSPQDSSSGIGTKGQPAKKRLNFGQDSE 180  
DB 121 AKKRVLEPLGLVEEVPKTPAGKRPVEHSPVEPDSSSGTGKAGQPAKKRLNFGQDAD 180  
QY 181 SVDPQPLCEPPATPAAGVPTTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVI 240  
DB 181 SVDPQPLGQPPAAGSLGTNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVI 240  
QY 241 TTSTRTWALPTYNHLYKQISSASTGASNDNHYFGYSTPWGDFDNRFFCHFSPRDWORL 300  
DB 241 TTSTRTWALPTYNHLYKQISSQS-GASNDNHYFGYSTPWGDFDNRFFCHFSPRDWORL 299  
QY 301 INNNNGFRKRLNFKLNFQVKEVTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQ 360  
DB 300 INNNNGFRKRLNFKLNFQVKEVTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQ 359  
QY 361 GCLPPFPADVFMIPQYGYLTNNGSQVGRSSFYCLEYPPSQMLRTGNFTSYTFEEVP 420  
DB 360 GCLPPFPADVFMIPQYGYLTNNGSQVGRSSFYCLEYPPSQMLRTGNFTSYTFEDVP 419  
QY 421 FHSSVAHSQSLDRMLNPLIDQVLYLNRTONQSGSAQNKDILLFSRGSFAGMSVQPKWLP 480  
DB 420 FHSSVAHSQSLDRMLNPLIDQVLYLNRTONQSGSAQNKDILLFSRGSFAGMSVQPKWLP 479  
QY 481 GPCYRQORYSKTKTNNNSNFTWTGASKYNLANGRESIINPGTAMASHKDDDEKFFPMGCV 540  
DB 480 GPCYRQORYSKTSADNNSEYSGTATKYLNGRSLVNPGPAMASHKDDDEKFFPQSGV 539  
QY 541 MIFGKESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFOSSTDPATGDVHAMG 600  
DB 540 LIFGKQSEKTNVDIEKWTIDEEIRTTN-VATEQYGSVSTNLQRNQAATAADVNTQG 598  
QY 601 ALPGMWQDRDYYLQGPVIAKIPHTDGHFHPSPLMGGFGLKHPPPQIILIKNTVPANPPA 660  
DB 599 VLPGMWQDRDYYLQGPVIAKIPHTDGHFHPSPLMGGFGLKHPPPQIILIKNTVPANPST 658  
QY 661 EFSATKFSFITYOSTGQVSVETLEWELQKNSKRWNPVEQYTSNVAKSANVDFTVDNNGL 720  
DB 659 TFSAAKFSFITYOSTGQVSVETLEWELQKNSKRWNPVEQYTSNVAKSANVDFTVDNNGV 718  
QY 721 YTEPRPIGTRYLTPRL 736  
DB 719 YSEPRPIGTRYLTPRL 734

## RESULT 10

AAV71168  
ID AAV71168 standard; Protein; 599 AA.

XX AC AAV71168;

XX AC AAV71168;

DT 08-SEP-2000 (first entry)

DE Adeno-associated virus serotype 1 capsid protein VP2.

XX Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;

KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
XX vaccine; transgene; VP2.

OS Adeno associated virus serotype 1.

XX W0200028061-A2.

XX 18-MAY-2000.

XX 02-NOV-1999; 99MO-US25694.

XX 05-NOV-1998; 98US-0107114.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Willson JM, Xiao W;

XX WPI; 2000-376571/32.

XX N-PSDB; AAD00772; AAD00778.

XX Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host -

PS Claim 7; Page 93-95; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP2 which is  
CC useful in the production of recombinant viral vector for gene delivery.

SQ Sequence 599 AA;

Query Match 81.5%; Score 3251; DB 21; Length 599;  
Best Local Similarity 100.0%; Pred. No. 1.7e-252;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 TAPGKRPVEQSPQEPDSSSGIGTKGQPAKKRLNFGQDSESVDPQPLGEPATPAA 197

DB 1 TAPGKRPVEQSPQEPDSSSGIGTKGQPAKKRLNFGQDSESVDPQPLGEPATPAA 60

QY 198 VGPTTMASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTWALPTYNHLY 257

DB 61 VGPTTMASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTWALPTYNHLY 120

QY 258 KOISSASTGASNDNHYFGYSTPWGDFDNRFFCHFSPRDWORLNNNGWFRKRLNFKLP 317

DB 121 KOISSASTGASNDNHYFGYSTPWGDFDNRFFCHFSPRDWORLNNNGWFRKRLNFKLP 180

QY 318 NIQKVEVTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQCLPPFPADVFMIPQY 377

DB 181 NIQKVEVTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQCLPPFPADVFMIPQY 240

QY 378 YLTNNGSQVGRSSFYCLEYPPSQMLRTGNFTSYTFEEVPFHSSVAHSQSLDRMLNP 437

DB 241 YLTNNGSQVGRSSFYCLEYPPSQMLRTGNFTSYTFEEVPFHSSVAHSQSLDRMLNP 300

QY 438 LIDQVLYLNRTONQSGSAQNKDILLFSRGSFAGMSVQPKWLPFCYQQRVSKTKTDNN 497

DB 301 LIDQVLYLNRTONQSGSAQNKDILLFSRGSFAGMSVQPKWLPFCYQQRVSKTKTDNN 360

QY 498 NSNFTWTGASKYNLANGRESIINPGTAMASHKDDDEKFFPMGCVMI FGKESAGASNTALDN 557

DB 361 NSNFTWTGASKYNLANGRESIINPGTAMASHKDDDEKFFPMGCVMI FGKESAGASNTALDN 420

QY 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 617  
|||  
DB 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
|||  
QY 618 IWAKIPHTDGHFHPSPLMGGFGLKNPPQIILINKTPVPANPPAEFSATKFAFITYSTG 677  
|||  
DB 481 IWAKIPHTDGHFHPSPLMGGFGLKNPPQIILINKTPVPANPPAEFSATKFAFITYSTG 540  
|||  
QY 678 QVSVLEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 736  
|||  
DB 541 QVSVLEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 599  
|||  
RESULT 11  
AAV71169  
ID AAY71169 standard; Protein; 534 AA.  
XX  
AC AAY71169;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 capsid protein VP3.  
XX  
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP3.  
XX  
OS Adeno associated virus serotype 1.  
XX  
PN WO200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PP 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
XX WPI; 2000-376571/32.  
DR N-PSDB; AAD00772, AAD00779.  
XX  
PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host  
XX  
PS Claim 7; Page 99-101; 108pp; English.  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP3 which is  
CC useful in the production of recombinant viral vector for gene delivery.  
XX  
SQ Sequence 534 AA;  
Query Match 72.9%; Score 2906; DB 21; Length 534;  
Best Local Similarity 100.0%; Pred. No. 8e-225;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 203 MASGGAPMADNNEGAGVGNAGNWHCDSTWLGDRVITTRTWTALPTYNHLYKQISS 262  
|||||

DB 1 MASGGAPMADNNEGAGVGNAGNWHCDSTWLGDRVITTRTWTALPTYNHLYKQISS 60  
QY 263 ASTGASNDNHFGYSTPMGYFDFNRHCHFSRDPQWRLNNNWGPRPRKLFKLFNIQVK 322  
|||  
DB 61 ASTGASNDNHFGYSTPMGYFDFNRHCHFSRDPQWRLNNNWGPRPRKLFKLFNIQVK 120  
|||  
QY 323 EVTTNDGVTTIANLNTSTVQVFSDEYQLPVVLGSAHQCLPPPPADVFMIPOYGYLTIN 382  
|||  
DB 121 EVTTNDGVTTIANLNTSTVQVFSDEYQLPVVLGSAHQCLPPPPADVFMIPOYGYLTIN 180  
|||  
QY 383 NGSQAVGRSSSYCYLEYPPSQMLRTGNFTFSTYTPPEVPHSSYAHQSLSLDRMLNPLIDQY 442  
|||  
DB 181 NGSQAVGRSSSYCYLEYPPSQMLRTGNFTFSTYTPPEVPHSSYAHQSLSLDRMLNPLIDQY 240  
|||  
QY 443 LYLNRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWLPGCYRQORVSKTKTDNNNSNFT 502  
|||  
DB 241 LYLNRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWLPGCYRQORVSKTKTDNNNSNFT 300  
|||  
QY 503 WTGASKYNLNGRESIINPGTAMASHKDDKDEFFPMGSMYIFGKESAGASNTALDNVMTD 562  
|||  
DB 301 WTGASKYNLNGRESIINPGTAMASHKDDKDEFFPMGSMYIFGKESAGASNTALDNVMTD 360  
|||  
QY 563 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 622  
|||  
DB 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 420  
|||  
QY 623 PHTDGHFHPSPLMGGFGLKNPPQIILINKTPVPANPPAEFSATKFAFITYSTQGVSV 682  
|||  
DB 421 PHTDGHFHPSPLMGGFGLKNPPQIILINKTPVPANPPAEFSATKFAFITYSTQGVSV 480  
|||  
QY 683 IEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 736  
|||  
DB 481 IEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 534  
|||  
RESULT 12  
AAG65793  
ID AAG65793 standard; Protein; 598 AA.  
XX  
AC AAG65793;  
XX  
DT 11-FEB-2002 (first entry)  
XX  
DE Adeno-associated virus 2 (AAV-2) major coat protein VP2.  
XX  
KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
KW inverted terminal repeat; neurotropic; neuroprotective; antianemic; ITR;  
KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
KW major coat protein; AAV-2; VP2.  
XX  
OS Adeno-associated virus 2.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1 /note= "encoded by ACG"  
FT  
PN WO200168888-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 13-MAR-2001; 2001WO-US07927.  
XX  
PR 14-MAR-2000; 2000US-189110P.  
XX  
PA (NEUR-) NEUROLOGIX INC.  
XX  
PI Xiao W, During MJ;  
XX  
DR WPI; 2001-596912/67.  
DR N-PSDB; AAI66974.  
XX  
PT Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a

chimeric capsid having one non-native amino acid sequence and a desired transgene

Disclosure, Page 51; 53pp; English.

The invention provides a recombinant viral vector (RVV) comprising a chimeric capsid (I) having at least one non-native amino acid sequence, derived from a capsid protein domain of parvovirus (II), a virus (III), or their combination, and a transgene flanked 5' and 3' by inverted terminal repeat (ITR) sequences, derived from (II), (III), or their combination. The RVV is useful for improving gene therapy in a subject with a disorder, and for increasing the efficiency of entry into a cell, which involves producing (I) encapsulating a viral vector, and contacting a cell with RVV having (I) such that (I) binds to an attachment site on the cell surface and permits the vector to enter the cell efficiently. A pharmaceutical composition comprising RVV with (I) containing a transgene sequence associated with a disease or a disorder such that expression of the transgene would result in amelioration of the disease or disorder such as inherited neurological and metabolic diseases e.g. lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polyneuropathy, Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic fibrosis, diabetes, diseases associated with hormone deficiencies, retinoblastoma and various types of neoplastic cells which include tumours especially central nervous system tumours, neoplasms, carcinomas, sarcomas, leukemias and lymphoma. The present sequence represents the adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences are used in the construction of a chimeric vector.

Sequence 598 AA;

```
Query Match          69.2%; Score 2759.5; DB 22; Length 598;
Best Local Similarity 82.8%; Pred. No. 5.8e-213;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 139 APGKKRPVEQSPQEPDSSSGIGTKGQAPKRLNFGQTGDSSEVPDPQLGEPPTAAV 198
DB 2 APGKKRPVEHSPVEPDSSSGTGKAGQAPKRLNFGQTGDSVDPQLGQPPAASGL 61
QY 199 GPTTASGAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTSRTWALPTNNHLYK 258
DB 62 GTNTWATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVITTSRTWALPTNNHLYK 121
QY 259 QISSASTGASNDNHYFGYSTPMGYDFNRFCHFPSPDWQRLNINNWGPRKRLNPKLN 318
DB 301 IDOYLILNRTQNSGSAQNKDLFSGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNN 498
QY 499 SNTFTWTCASKNYLNGRESIINPGTVMASHKDDKFPKMSGVMI FGESAGASNTALDNV 558
DB 361 SEYSWTGATKYLHNGRDSLVNPGPAMASHKDDSEKFPQSGVLI FGQSEKNTNVDIEK 420
QY 559 MITDEBEIKATNPVATRFQTVAVNFQSSSTDPAITGDVHAMGALPGVMWDRDYLQGP 618
DB 421 MITDEBEIRTNVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMWMDRDVYLQGP 480
QY 619 WAKIPHTDGHFSPMLMGFGFLKNPPQIILKNTVPANPPAPSFATKFSFTQVSTGQ 678
DB 481 WAKIPHTDGHFSPMLMGFGFLKNPPQIILKNTVPANPPSTTFSAAKFSFTQVSTGQ 540
QY 679 VSVIEIWELOKENSKRWNPEVQVTSNYAKSANVDFTVDNNGLYTERPEPIGTRYLTRPL 736
```

Db 541 VSVIEIWELOKENSKRWNPEIQVTSNYKSNVNDFTVDNNGVYSEPRPIGTRYLTRNL 598

RESULT 13

AAVS1509

ID AAMS1509 standard; protein; 598 AA.

AC AAMS1509;

DT 02-JAN-2002 (first entry)

DE Adeno-associated virus VP2 capsid protein.

KW Adeno-associated virus; AAV; VP2; capsid; virus-like particle;

OS Adeno associated virus.

PN JP2001169777-A.

PD 26-JUN-2001.

PF 30-JUL-1999; 99JP-0249140.

PR 30-JUL-1999; 99JP-0249140.

PA (HAND/) HANDA H.

DR WPI; 2001-599854/68.

XX New virus-like particles from VP3 capsid protein of adeno-associated virus, comprise a peptide containing a nucleus-shifting signal connected to its N-terminal -

XX Claim 1; Page 14-16; 33pp; Japanese.

XX The present sequence is provided in a specification relating to a virus-like particle-forming protein, and to a peptide containing a nucleus-shifting signal at its N-terminus that can form virus-like particles by shifting to the nucleus of the animal cell in which it is expressed. The method is used for forming virus-like particles from the VP3 capsid protein of adeno-associated virus (AAV).

Sequence 598 AA;

```
Query Match          69.2%; Score 2759.5; DB 22; Length 598;
Best Local Similarity 82.8%; Pred. No. 5.8e-213;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 139 APGKKRPVEQSPQEPDSSSGIGTKGQAPKRLNFGQTGDSSEVPDPQLGEPPTAAV 198
DB 2 APGKKRPVEHSPVEPDSSSGTGKAGQAPKRLNFGQTGDSVDPQLGQPPAASGL 61
QY 199 GPTTASGAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTSRTWALPTNNHLYK 258
DB 62 GTNTWATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVITTSRTWALPTNNHLYK 121
QY 259 QISSASTGASNDNHYFGYSTPMGYDFNRFCHFPSPDWQRLNINNWGPRKRLNPKLN 318
DB 122 QISSQS-GASNDNHYFGYSTPMGYDFNRFCHFPSPDWQRLNINNWGPRKRLNPKLN 180
QY 319 IOVKEVTTNDGVTITANNLTSTVQVFTDSEYQLPYVLGSAHOGCLPPFPADVPMVPOYGY 378
DB 181 IOVKEVTTNDGVTITANNLTSTVQVFTDSEYQLPYVLGSAHOGCLPPFPADVPMVPOYGY 240
QY 379 LTLNNGSAVGRSSFYCLEYFPQMLRTGNNTFFSYTFEVPFHSSYAHQSQSLDRLMNP 438
DB 241 LTLNNGSAVGRSSFYCLEYFPQMLRTGNNTFFSYTFEVPFHSSYAHQSQSLDRLMNP 300
QY 439 IDOYLILNRTQNSGSAQNKDLFSGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNN 498
DB 301 IDOYLILNRTQNSGSAQNKDLFSGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNN 360
```



XX		Chiorini JA, Kotin RM, Safer B;
XX	PI	WPI; 1998-207403/18.
DR	DR	N-PSDB; AAV21648, AAV21650.
XX		
XX		Adeno-associated virus 4 based vectors - used to transduce erythroid progenitor cells for treatment of e.g. cancer and other genetic diseases
PT		
XX		
PS		Claim 34; Page 54-55; 80pp; English.
XX		
CC		This polypeptide comprises the adeno associated virus 4 (AAV4) VP1 capsid protein that is encoded by open reading frame 2 (see AAV21650) of the AAV4 genome (see also AAV21648). The invention also provides AAV4 VP2 (see AAM46313) and VP3 (see AAM46314), also encoded by the AAV4 genome (see also AAV21657-58). The invention provides recombinant vectors and viral particles based on AAV4 that may be useful for transducing erythroid progenitor cells for the treatment of e.g. cancer and genetic diseases which can be corrected by bone marrow transplants using matched donors. A claimed method of delivering a nucleic acid to a subject (including a subject with antibodies to AAV2) comprises administering to a cell from the subject an AAV4 particle comprising the nucleic acid inserted between a pair of AAV inverted terminal repeats (see AAV21651 and AAV21659), and returning the cell to the subject.
XX	SQ	Sequence 734 AA;
		Query Match 62.3%; Score 2486.5; DB 19; Length 734;
		Best Local Similarity 63.5%; Pred. No. 7,1e-191;
		Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;
Qy	4	DGYLPDWLEDNLSEIREMWDLPGAPKPKANQQDDGRGLVLPGYKYLGFNGLDKGE 63
Db	3	DGYLPDWLEDNLSEGVEEMWALQCPGAPKPKAQHQDNARGLVLPGYKYLFGNGLDKGE 62
Qy	64	PVNADAAAALBHDKAYDOOLKAGDNPYLRYNHDAEFQERLOEDTSPCGNLGRAVFOAKK 123
Db	63	PVNADAAAALBHDKAYDOOLKAGDNPYLKYNHDAEFOORIQGDTSPCGNLGRAVFOAKK 122
Qy	124	RVLPLGLIVEGAKTAPGCKRPVEOSPOBPSSSGIGTKTGQPAPKKRLNF-QGTGDSRV 182
Db	123	RVLPLGLIVEAGETA PGCKRPLIESPQDPSSSTGIGKKQPAKKLVFDEETGAGDGP 182
Qy	183	PDPOPGLBPAPTAAVGPTTWASGGGA PMADNEGADVGNASGNHCDSTWLGDRTVT 242
Db	183	PEGSTSG-----AMSDDSEMRAAAGAAVEGGQGADVGNASGDWHCDSTWSEGHVTT 236
Qy	243	STRTWALPTYNNHLVKYI SSASTGASNHNHYFGYSTPMGYPDFNRFCHFSPRDMQRLIN 302
Db	237	STRTWALPTYNNHLVKRI GE-----SLQNTYNGFTSPMGYDFDNRFCHFSPRDMQRLIN 292
Qy	303	NNWGFRPKRLNFKLFNI QVKEVYTNDGVTTIANLLTSTVFSDSEYQLPYVLGSAHQGC 362
Db	293	NNWGRBRKAMRVKI FNI QVKEVTTNGETTVANNLTSTVQIFADSSYELPYVMDAQSGS 352
Qy	363	LPPPPADVFMIPQYY----LTNNGSQAVGRSSFICYLFYFPSQMLRTGNFFSYTFPEV 419
Db	353	LPPFNDVFMVPQYGYCGLVLTGNTSQOQTDRNAFYCLEYFPPSQMLRTGNFFITYSFKY 412
Qy	420	PHSSHVAHSQSILDRLMNPLIDQLYLYLNARTQN---QSGSAQNKDLLLSRSGSPAGMSVQP 475
Db	413	PFHSHVAHSQSILDRLMNPLIDQLYLGLOSITTTGGTTLNAGTATTN---FTKRPTWFNSFK 469
Qy	476	KNWLPGPCVRQQRVSKTCTDNNNSFTWTGAS---KY-----NLNGRESLIINPGTAMASHK 528
Db	470	KNWLPGPSIKQGFPSKTA--NQNYKIPATGSDSLIKYETHSTLDGRWSALTGPPMATAG 527
Qy	529	DDEKFFPMGSWMIVFGKSAGASNTALDNVMTIDEEETKATNPVATERPGTVVNFQSSS 588
Db	528	PADSK-FSNSQIIIFAGPQONGNTATVPGLTIPTSEEELAATNATQTDMMGNLPGDGQSN 586

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:00:42 ; Search time 17.3269 Seconds  
(without alignments)  
4084.983 Million cell updates/sec

Title: US-09-807-802A-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWLENDLSEGR.....NNGLYTEPRPIGTVLTRL 736  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2177.5	54.6	504	1 VCPV3A	coat protein - ade
2	2168.5	54.4	732	2 S52210	coat protein VP1
3	798.5	20.0	673	1 VCPVB5	coat protein VP1
4	600.5	15.1	781	1 VCPV19	coat protein VP1
5	482	12.1	729	1 VCPVNA	coat protein VP1
6	482	12.1	729	1 A60006	coat protein VP1
7	444.5	11.1	727	1 VCPV1F	coat protein VP1
8	439.5	11.0	722	1 VCPVME	coat protein VP1
9	431.5	10.8	727	1 VCPVFP	coat protein VP1
10	426	10.7	723	1 VCPVFP	coat protein VP1
11	424.5	10.6	718	1 VCPVIM	coat protein VP1
12	421.5	10.6	748	1 VCPVCP	coat protein VP1
13	407.5	10.2	737	1 VCPVCD	coat protein VP1
14	404	10.1	722	1 VCPV2	coat protein VP1
15	385.5	9.7	722	1 VCPVCN	coat protein VP1
16	377	9.5	716	1 VCPV2M	coat protein VP1
17	234	5.9	587	1 B44276	coat protein VP1
18	199.5	5.0	810	2 A44054	orf1 protein - Jun
19	198	5.0	584	2 A49594	capsid protein VP2
20	191	4.8	702	1 VCPVAP	coat protein VP1
21	143.5	3.6	648	2 S50856	whn protein - rat
22	139	3.5	931	2 T49710	related to glucan
23	131.5	3.3	1446	1 A45344	immediate-early pr
24	130.5	3.3	1145	2 T18235	transcription acti
25	130	3.3	1338	2 T30565	MAP kinase kinase
26	129	3.2	1072	2 A86827	hypothetical prote
27	124.5	3.1	2493	2 A55481	adenylate cyclase
28	124	3.1	1296	2 C82521	hemolysin-type cal
29	124	3.1	3078	2 T28432	variant-specific s

ALIGNMENTS

RESULT 1  
VCPV3A

coat protein - adeno-associated virus type 2  
C:Species: adeno-associated virus type 2  
C:Date: 05-Apr-1983 #sequence\_revision 05-Apr-1983 #text\_change 16-Jul-1999  
C:Accession: A03698

R:Srivastava, A.; Lusby, E.W.; Berns, K.I.  
J. Virol. 45, 555-564, 1983

A:Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.  
A:Reference number: A03694; MUID:83164299; PMID:6300419  
A:Accession: A03698

A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-504 <SRI>

A:Cross-references: EMBL:J01901; NID:G209616; PIDN:AAA42376.1; PID:G209621; EMBL:M12405;  
C:Superfamily: adeno-associated virus coat protein  
C:Keywords: coat protein

Query Match 54.6%; Score 2177.5; DB 1; Length 504;  
Best Local Similarity 82.5%; Pred. No. 1.3e-134;  
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;

QY	203	MASGGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYQISS	262
DB	1	MATGSGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYQISS	60
QY	263	ASTGASNDNHFGYSTPMGYDFNRFCHFSRDMORLINNNWGRPKRLNFKLFNIQVK	322
DB	61	QS-GASNDNHFGYSTPMGYDFNRFCHFSRDMORLINNNWGRPKRLNFKLFNIQVK	119
QY	323	EYTTNDGVTTIANNLTSTVQFSDSEYQLPVYVLSAHQGLPPFPADVFMIPOYGYLTIN	382
DB	120	EYQNDGTTTIANNLTSTVQFSDSEYQLPVYVLSAHQGLPPFPADVFMIPOYGYLTIN	179
QY	383	NGSQAVGRSSFCYCLEYFPFQMLRTGNNFTFSYTFPEVPHSSYAHQSILDRMLNPLIDQY	442
DB	180	NGSQAVGRSSFCYCLEYFPFQMLRTGNNFTFSYTFPEVPHSSYAHQSILDRMLNPLIDQY	239
QY	443	LYLNTQNSGSAQNKDILFSRGSFAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNFT	502
DB	240	LYLNTQNSGSAQNKDILFSRGSFAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNFT	299
QY	503	WTGASKYNLNGRESIINPGTAMASHKDDKDFPFMSGVMIFGKESAGASNTALDNVMTD	562
DB	300	WTGASKYNLNGRESIINPGTAMASHKDDKDFPFMSGVMIFGKESAGASNTALDNVMTD	357
QY	563	EEIEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKI	622
DB	358	EEIEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKI	417
QY	623	PHTDGHFSPMLMGFGFLKNPPQILIKNTVPANPPAEFSATKPFASITQSTG	677

probable TonB-depe  
toxin-like outer m  
enamelin precursor  
valine-tRNA ligase  
protein F2K11.10  
nuclear pore compl  
type VII collagen  
hypothetical serin  
hypothetical prote  
IGA-specific serin  
probable membrane  
transcription fact  
toxin-like outer m  
probable transakeo  
myrosinase-binding  
nuclear pore prote

Db 418 PHTDGHFHPSPLMGGFLGKPPQILIKNTVPANPSTTTSAAKFASFIQYSTG 472

RESULT 2

S52210

coat protein VP1 - muscovy duck parvovirus

N:Alternate names: VP1 protein

C:Species: muscovy duck parvovirus

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-2000

C:Accession: S52210

R:Zadori, Z.; Erdei, J.; Nagy, J.; Kisari, J.

submitted to the EMBL Data Library, September 1993

A:Reference number: S52209

A:Accession: S52210

A:Molecule type: DNA

A:Residues: 1-732 <ZAD>

A:Cross-references: EMBL:X75093; NID:G609091; PIDN:CAAS2984.1; PID:G609093

A:Experimental source: strain FM

C:Genetics:

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

Query Match 54.4%; Score 2168.5; DB 2; Length 732;

Best Local Similarity 55.2%; Pred. No. 8.7e-134;

Matches 418; Conservative 90; Mismatches 186; Indels 63; Gaps 13;

QY 9 DWLEDNLSEGIREWDLKPGAPKPKANQKQ-----DQGLGLVLPQYKYLPGFN 57

Db 10 DYTE-----TAAASRWHLKAGAPKPKSNQOSQSVSTRPKQRKNNRGFLVPGYKVGPGN 65

QY 58 GLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQRLQEDTSFGNGLGRA 117

Db 66 GLDKGPPVNAKDSVALEHDKAYDQQLKAGDNPYIKFKHADQEFIDNLQDTSFGNGLGKA 125

QY 118 VFQAKRVLPLGLVEEGANTAPGKRVPQSPQDSSSGIGTKGQOPAKRLNFGQTG 177

Db 126 VFQAKRIILEPLGLVEEPVNTAPAK-----SSG-KLTDHDPVVKPKLSE-- 170

QY 178 DSESPDPQPLG-----EPATPAAVGPTTMSAGGAPMADNNEGADGVGNAG 226

Db 171 --ENSPSPNSGGEASAAATESEPVAP-----NMAEGSGGAMGDSAGGADGVGNAG 222

QY 227 NWHCDSTWLGDRVITTTSTWALPTNNHLYKQISSASTGASNDNHFYGYSTPMGYFDNF 286

Db 223 NWHCDSQWLGDTVITKTITWVLPSTNNHMYQAITSNTPDNSN-TQYAGYSTPMGYFDNF 281

QY 287 RFCHFSPRDWORLNNHNGFPRKLNFKLFIQVKEVTTNDGVTTIANLSTVQVFS 346

Db 282 RFCHFSPRDWORLNNHNGFPRKLNFKLFIQVKEVTTNDGVTTIANLSTVQVFS 341

QY 347 SEVQLPYVLGSAHOGCLPPFPADVPMIPOYGYLTN---NGSQAVGRSFFCYCLEYFPSSQ 403

Db 342 NEHQLPYVLGSAHOGCLPPFPADVPMIPOYGYLTN---NGSQAVGRSFFCYCLEYFPSSQ 401

QY 404 LRTGNNFTSYTFEEVPHSSYAHOSQLRLNPLIDQYLYLNRNTQSGSAQNKLDF 463

Db 402 LRTGNNFTSYTFEEVPHSSYAHOSQLRLNPLIDQYLYLNRNTQSGSAQNKLDF 455

QY 464 SRGSPAGMSVQPKNMLPGCYRQORVSKTK---TDNNNSNFTWTGASKYNLNGRESIINPG 521

Db 456 KXAVGAFGAGGRNMLPGKLLDQVRAYSGGTDYANWSIWSKGNKVFLDKREYLLQPG 515

QY 522 TAMASHKDEDKFFPMSPGVMIFGKE--SAGASNTALDNVMTDEEIKATNPVATERPGT 579

Db 516 PVATHTTEQASVPAQNIIGAKDYRSGSTLAGISDIWTDQEIAPNTGVGWRPYGL 575

QY 580 VAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPLMGGFG 639

Db 576 TVTNEQNTTAPTNAEVLGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPLMGGFG 635

QY 640 LKNPPQIILIKNTVPANPASTKFAFQYSTGVSVEIEHLEKENSKRWNPEV 699

Db 636 LHNPPPPQVFIKNTVPADPPPLEYVKNQNSYITQYSTGCTQCTVEMVWELRKENSKRWNPEI 695

QY 700 QYTSNTYAKSANVDFTVDNGLYTEPRPIGRYLRPL 736

Db 696 QYTSNFGNISTWFAFNETGCGYVEDRLIGRILYLTQNL 732

RESULT 3

VCPVB5

N:Contains: coat protein VP1 - bovine parvovirus

C:Species: bovine parvovirus

C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999

C:Accession: A26104

R:Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.

J. Virol. 60, 1085-1097, 1986

A:Title: Complete nucleotide sequence and genome organization of bovine parvovirus.

A:Reference number: A26104; MUID:87061184; PMID:3783814

A:Accession: A26104

A:Molecule type: DNA

A:Residues: 1-673 <CHE>

A:Cross-references: EMBL:M14363; NID:G333454; PIDN:AAB59847.1; PID:G808805

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

F:138-673/Product: coat protein VP2 #status predicted <VP2>

Query Match 20.0%; Score 798.5; DB 1; Length 673;

Best Local Similarity 28.6%; Pred. No. 2.5e-44;

Matches 214; Conservative 105; Mismatches 275; Indels 155; Gaps 21;

QY 30 PKPKANQKQDDGRGLVLPQYKYLGPFPNGLDKGPVNAADAAALEHDKAYDQQLKAGDNP 89

Db 3 PTNKANSKK-----GLTLPQYNYLGFNSLFAQAPVNAKADAAARKHDFGYSDDLKEGKNP 57

QY 90 YLRVNHADAEFQRLQEDTSFGNGLGRAVQAKRVLPLGLVEEGAKTAPGKRVPQES 149

Db 58 YLYFNTDQNLIDELKDDTSFGGKLARGVFOIKALAPAL-----PQTSKGGDRALKRK 111

QY 150 PQEPDSSSGIGTKQOPAKKRLNFGQCDSESVPDPQPLGEP--PATPAAVGPTTMSAGG 207

Db 112 LYFARSKGAKANREPAPSTSNQNNMEVNDIPNDEAGNQPIELATRSVVGSGSVGGG 171

QY 208 GAPMADNNEGADGVGNASGNHCDSTWLGDRVITTTSTWALPTNNHLYKQISSASTGA 267

Db 172 -----RQSGVGYSTGWTGTFISENIWVTKTRQFICDKNGHLYKS-EVLNTGD 222

QY 268 SNDNHFYGYSTPMGYFDNFPHCHFSPRDWORLNNHNGFPRKLNFKLFIQVKEVTTN 327

Db 223 TAHRQY-AITTPWSYFNFQYSHFSFNDWQHLVNDYERFAPKAMIVRVYMLQIKQIMTD 281

QY 328 DGVTTIANN-LTSTVQVFSSEYQLPYVLGSAHOGCLPPFPADVPMIPOYGYLTN----- 381

Db 282 GAMGTVYNNDLTAGMHIFCDGDHRYPVYQHPWDQCPPELNSIWELPQYAYIPAPISV 341

QY 382 -NGSQAVGRS-----SFYCLEYFPSSQMLRTGNNTFTSYTFEEVPHSSYAHOSQLRLN 435

Db 342 DNNTTNTVEEHLKGVPLYLENSDHEVLRNG-----RIV 376

QY 436 NPLIDQYLYLNRNTQ-----QSGSAQNKLDFSRGSP-----AGMSVQPKNMLP 480

Db 377 RIYIOLMRLNRMDRKHQHIQHASDDVQSTGQKQKLLIQTQKQPKQRFQNAALRTSNMMS 436

QY 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDKFFPMSPG 540

Db 437 GP-----GIARGTHNATLQTSAGALVTWVTNGAD-----VSGV 470

QY 541 M-----IPGKE-----SAGASNTALDNVMTDEEIKATNPVATE 575

Db 471 RAVRVGYSTDPYGGQPPESDLLRLYSASAAGQQNPILN-----AARH 516

QY 576 RFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPLM 635

Db 517 TTTREARTKLTGSGNAGDGYKEWMLPNQMWDSAPISRYNPIWVKVPRVNRKTLTLDQD 576

```
QY 636 GGGLKNPPQILIK--NTPVPANPPAEFSATKFAFSTOYSTGVQSVSEIWELOKENS 693
Db 577 GSIPMGHPPTGIFIKLARIPVPGND-----SFLNIYVTGVQSVCEVWEVERGTK 627
QY 694 RMNPEVOYTSNYAKSANVD--FTVDNGLY 721
Db 628 NWRPEYMHS---ATNMSVDAYTINNAGVY 653

RESULT 4
VCPV19
coat protein VP1 - parvovirus B19 (strain Au)
C:Species: parvovirus B19
A:Note: host Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: A24299
R:Shade, R.O.; Blundell, M.C.; Cotmore, S.F.; Tattersall, P.; Astell, C.R.
J. Virol. 58, 921-936, 1986
A:Title: Nucleotide sequence and genome organization of human parvovirus B19 isolated from
A:Reference number: A24299; MUID:86200451; PMID:3701931
A:Accession: A24299
A:Molecule type: DNA
A:Residues: 1-781 <SHA>
A:Cross-references: EMBL:M13178; NID:G333375; PIDN:AAA66867.1; PID:G333377
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein

Query Match 15.1%; Score 600.5; DB 1; Length 781;
Best Local Similarity 26.8%; Pred. No. 2.7e-31;
Matches 182; Conservative 95; Mismatches 300; Indels 107; Gaps 20;

QY 47 LPGYKYLPGPNGLDKGEPVNAADAAAEHDKAYDQQLKAGDNPYLRYNHADAEEFORLOE 106
Db 125 LPGNTVYGPNEGLOAGPQSAVDSAAIHDPRYSQAKLGINPTHTWVADELLKNIKN 184
QY 107 DTSFGNGLGRAVFOAKKRVLEPLGLVEGAKTAPGKKRPVQSPQBPDPSSSGIGTKGQOP 166
Db 185 ETGFQAVQVVDYF-----TLKGAAPVNHQ-----GSLPEVP 217
QY 167 AKRLNFGQGDSESPDPQPLGEPATPAAVGPTTMASSGGGAPMADNNEGADGVGNASG 226
Db 218 AYNA-----SEKPSMTSVNSAEASTGA-----GGG-----GNSVKS 250
QY 227 NMHCDSTWLGDRTVITSTRTWALPTNNHLYKQISSASTCASND-----NHFGY 276
Db 251 MWSEGAFTSANSVCTFSRQLPYDPEHYKYVSPAASCHNASCKEAKVCTISIMGY 310
QY 277 STPWGTFDPRFCHFSPRDWQRLNNMNGFRPKRLNFKLFNTQVKEVT--TNDGVTTIA 334
Db 311 STPWYLDNFALNLFPSLEFQHLIENYGSIAPDALTVTISETIAVKDVTDKTGGV-QVT 369
QY 335 NNLTSVTQVFSDEYQLPYVLGSAHQCLPPFPADVPMPQCYLTLNN-GSAVG----389
Db 370 DSTTGLRCLMLVDREYKPYVLVGQGDQLAPELPIWTFPPQYALTVGDVNTQGISGDSK 429
QY 390 -----RSSFYCLEYFYSQMLRTGNFTSYTFEEVPFHSYAHSSQSLDRMLNPLIDQYLY 444
Db 430 KLASBESAFVLEHSSFQLLGTGTASMSYKFPVPENLEGCSQHFYEMYNPL----YGS 486
QY 445 YLNRTONQSSAQNKOLLFSRGSAGMSVQPKWLPGPCYQQRVSKTKTDNNNSFTWT 504
Db 487 RLGVDPDTLGGDPKFRSL-----THEDIAIQPMFPGPLVNSVSTKEGDSSTNAGKALT 541
QY 505 GASKYNLNGRESIINPG-TAMASHKDEDEKFFPMSGVMIFGKESAGASNTALDNV-----558
Db 542 GLSTGTSQNTRIELRPGVSPQYHHHDDTKYIVTGINAISHGQTYG---NAEDKEYQQGV 598
QY 559 --MITDEEEIKATNPVATERFGTVAVNFQSSSDPATGDVHAMGALPGVMWQDRDYVLOG 616
Db 599 GRFPNEKEQLKQLOGLNMHTY-----FPNKGTFQYTDQIE-RPLMVGSVWNRALHYES 651
QY 617 PIWAKIPHTDGHFHS-PLMGGFGLKNPPPPQIILIKNTPTVPANPPAEFSATKFASTQYS 675
```

```
Db 652 QLWSKIPNLDDSFQTKPFAALGGLGHQPPQPLFK--ILPQSGPIGGIKSMGITTLVQVA 709
QY 676 TQQSVSEIWEIWEI--QXENSKRWNP 698
Db 710 VGIIMTVMTFTKLGPRKATGRWNPQ 733

RESULT 5
VCPVNA
coat protein VP1 - porcine parvovirus (strain NADL-2)
N:Contains: coat protein VP2
C:Species: porcine parvovirus
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: B33743; D48472
R:Vasudevacharya, J.; Basak, S.; Srinivas, R.V.; Companys, R.W.
Virology 173, 368-377, 1989
A:Title: Nucleotide sequence analysis of the capsid genes and the right-hand terminal pa
A:Reference number: A33743; MUID:90085785; PMID:2596019
A:Accession: B33743
A:Molecule type: DNA
A:Residues: 1-729 <VAS>
A:Cross-references: GB:M32787; NID:G332983; PIDN:AAA46917.1; PID:G332985
R:Bergeron, J.; Menezes, J.; Tjissen, P.
Virology 197, 86-98, 1993
A:Title: Genomic organization and mapping of transcription and translation products of t
A:Reference number: A48472; MUID:94025614; PMID:8212598
A:Accession: D48472
A:Molecule type: DNA
A:Residues: 11-729 <BER>
A:Experimental source: NADL-2, ATCC VR-742
A:Note: sequence extracted from NCBI backbone (NCBIN:138789, NCBIP:138794)
C:Genetics: 10/1
A:Introns:
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein; glycoprotein
F:151-729/Product: coat protein VP2 #status predicted <VP2>
F:32,172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 12.1%; Score 482; DB 1; Length 729;
Best Local Similarity 25.5%; Pred. No. 1.3e-23;
Matches 197; Conservative 111; Mismatches 310; Indels 154; Gaps 35;

QY 43 RGLVLPGYKYLPGPNGLDKGEPVNAADAAAEHDKAYDQQLKAGDNPYLRYNHADAEEFOR 102
Db 9 RGLTLPGYKYLPGNSLQGEPTNPSDAAAEHDEAYDKYIKSGKNPYFPFAADEKFK 68
QY 103 RLQEDTSFGNGLGRAVFOAKKRVLEPLGLVEGAKTAPGKKRPVQSPQBPDPSSSGIGK- 161
Db 69 ETEHAKDYGGKIGHYFFRAKGAFAKPL---SETDSPTTSQOPEVRRSPRKHGPKGPKGR 125
QY 162 -----TGQOPAKKRLNFGQTDGDSSESPDPQPLGEPATPAAVGPTTMASSGGGAPMA 212
Db 126 PAPRIHIFINLAKKAKAGTSNTNSMSNEVQHNFINAGTEL-SATGNEGGGGGGG----181
QY 213 DNNEGADGVGNASG--NMHCDSTWLGD---RVITSTRTWALPTNNHLYKQIS--SAST 265
Db 182 -GGRGAGGVGVSTGTFTNNQTEFQYLGEGLVRITAHASRLIHLNMPHEHYTKRIHLNSES 240
QY 266 GAS-----NDNHVFGVSTPWGYFDNRFCHFSPRDWQRLNNMNGFRPKRLNFKLPNIQV 321
Db 241 GVAGQMVQDDAHTQVMVTPWLSLDANAWGVWFPADWQLISNNMTINLVSPQEIPNVVL 300
QY 322 KEVT---TNDGVTTIANLTSVQVFSDEYQLPYVLGSAHQCLPPFPADVPMPQCY 378
Db 301 KTIITESATSPPTKIYNNDLTASLWALDNTNLTPTPAAPRSETLGFYPMWLPKTPQYR 360
QY 379 L-----TLNNGSAQVGRS-----SFYCLE-YFPQMLRTGNF--TFSYTE 417
Db 361 YLSCIRNLNPPYTGSSQIITDSIQTGLHSDIMFYTIENAVPIHLRLTGDEBFTGIYHFD 420
QY 418 EVPFHSVAHSQSLDR-----LMNPLI---DQVLYINRTQNSQSGAQNKDLFSSRS 467
```



```
Db 421 TKPL--KLTHSWQNRSLGLPPKLLTEPTTBEGDQHPGTLPAANTRKGHYHOTINNSYTEAT 478
Qy 468 PAGMSVOPKNWLPQCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMA 525
Db 479 ----AIRP-----AQVGYNTPYMNFEYSNGGPF-----LTPIVPTADT 512
Qy 526 SHKDDKEDKPPPMGSMVIMFGKESAGASNTALD---NVMITDSEEEK--ATNP-----VAT 574
Db 513 QYNDDEPN-----GAIRFTMDYQHGHLTTSQELERYTFNFSQKGRAPK 557
Qy 575 ERFGTVA-VNFQSSS-----TDPATG--DVHMGALP-----GMVWQDRDYLQGIPIWA 620
Db 558 QQFNQAPLNTLNTNGTLLPSDPIGGKSNHFMNTLNTYGPLTALNNTAPVFPNGQIWD 617
Qy 621 KIPHTD--GHFHPSPLMGGFLK-NPPQIILIKNTVPANPAEFSA-TKPAFTIYOST 676
Db 618 KELDTDLKPLH--VTAPFVCKKNPPQQLFKIAP---NLTDNFNADSPQPRLIITYSN 671
Qy 677 QGVSVIEIWELOKENSKEWNPVQVTSNYAKSANVDFTVDNNGLYTEPRPIG 728
Db 672 FWMKGTLTFTAKMRSSNMNPIQHTT-----TAENIGNYI-PTNIG 712

RESULT 6
A60006
coat protein VP1 - porcine parvovirus (strain 90HS)
N:Contains: coat protein VP2
C:Species: porcine parvovirus
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C:Accession: A60006
R:Sakurai, M.; Nishimori, T.; Ushimi, C.; Nakajima, H.
Virus Res. 13, 79-86, 1989
A:Title: Nucleotide sequence of capsid protein gene of porcine parvovirus.
A:Reference number: A60006; MUID:89319168; PMID:2750278
A:Accession: A60006
A:Molecule type: DNA
A:Residues: 1-729 <SAK>
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein; glycoprotein
F:151-729/Product: coat protein VP2 #status predicted <VP2>
F:172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #status
Query Match 12.1%; Score 482; DB 1; Length 729;
Best Local Similarity 25.7%; Pred. No. 1.3e-23;
Matches 197; Conservative 112; Mismatches 314; Indels 144; Gaps 35;
Qy 43 RGLVLPGYKYLGPFGNLDKGEVPVNAADAAALSHDKAYDQCLKAGNPYLRYNHADAEFOE 102
Db 9 RGLTLPGYKYLGPFGNLDQGEPTNPSDAAAKEHDEAYDKYIKSGKNPYFYFSAADEKFIK 68
Qy 103 RLQEDTSFGGNLGRAVFOAKKRVLPGLGLVEEGAKTAPGKKRPVQSPQEPDSSSGIGK- 161
Db 69 ETEHAKDYGGKIGHYFFRA-KRAFRP--KLSETDPTTSQQPEVRSRKHPGSKPPGKR 125
Qy 162 -----TGQOPAKRLNFGQTDGSESVPDPQGLGEPPATPAAGVPTTMASGGGAPMA 212
Db 126 PAPRHIFINLAKKAKGTSTNSNSNSNVEQHNINAGTEL-SATGNSGGGGGGG--- 181
Qy 213 DNNEGADGVGNAGNWH--CDSTWLGD---RVITSTRTWALPTYNNHLYKOIS--SAST 265
Db 182 -GGRGAGGVGVTSGFNGNTEFOYLGGLVRITAKASRLIHLNMPHEHYTKRIHVNSES 240
Qy 266 GAS----NDNHYFGYSTPWGYDFNRFHCHFSRQWQRLNNWGRPKRLNFKLFNIQV 321
Db 241 GVAGQMVQDAHTQMTVTPWSLIDANAWGVFNPAWQQLISNNWTEINLVSPFEQEIFNVVL 300
Qy 322 KEVT---TNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPPPADVFMIPOQY 378
Db 301 KITESATSPPTKIYNNNDLTASLMVALDNTNLTPTPAAPRSETLGFYFWLPTKPTQVRY 360
Qy 379 L-----TLNNGSQAVGRS-----SFYCLE-YFPQMLRTGNF-TFSYTFE 417
Db 361 YLSCTRNLPNPPYTGSSQOITDSIQTLGHSIDIMFYTIENAVPIHLRLTGDEFSTGIYHPD 420
```

```
Qy 418 EVFHHSSVAHSQSGLDR-----LMNPLI--DQYLYLNRNTQNSGSAQNKDLLFSRGS 467
Db 421 TKPL--KLTHSWQNRSLGLPPKLLTEPTTBEGDQHPGTLPAANTRKGHYHOTINNSYTEAT 478
Qy 468 PAGMSVOPKNWLPQCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMA 525
Db 479 ----AIRP-----AQVGYNTPYMNFEYSNGGPF-----LTPIVPTADT 512
Qy 526 SHKDDKEDKPPPMGSMVIMFGKESAGASNTALDVMITDSEEEKATNP-----VATERFGT 579
Db 513 QYNDDEPN-----PAGAIRFTMDYQHGHLTTS-----SQELERYTFNFSQKGRAPKQGFNQ 562
Qy 580 VA-VNFQSSS-----TDPATG--DVHMGALP-----GMVWQDRDYLQGIPIWAKIPIHT 625
Db 563 QAPLNLNTNGTLLPSDPIGGKSNHFMNTLNTYGPLTALNNTAPVFPNGQIWDKELDT 622
Qy 626 D--GHFHPSPLMGGFLK-NPPQIILIKNTVPANPAEFSA-TKPAFTIYOSTGVSV 681
Db 623 DLKPLH--VTAPFVCKKNPPQQLFKIAP---NLTDNFNADSPQPRLIITYSNFWMKG 676
Qy 682 EIWELQKENSKEWNPVQVTSNYAKSANVDFTVDNNGLYTEPRPIG 728
Db 677 TLFTAKMRSSNMNPIQHTT-----TAENIGNYI-PTNIG 712

RESULT 7
VCPVIF
coat protein VP1 - feline panleukopenia virus
N:Contains: coat protein VP2
C:Species: feline panleukopenia virus, FPLV
C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A03701
R:Carlsson, J.; Rushlow, K.; Maxwell, I.; Maxwell, S.; Hahn, W.
J. Virol. 55, 574-587, 1985
A:Title: Cloning and sequence of DNA encoding structural proteins of the autonomous parvovirus
A:Reference number: A03697; MUID:85265017; PMID:2991581
A:Accession: A03701
A:Molecule type: DNA
A:Residues: 1-727 <CAR>
A:Cross-references: EMBL:M10824; NID:g333474; PIDN:AAA47161.1; PID:g333476
C:Genetics:
A:introns: 11/1
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein
F:144-727/Product: coat protein VP2 #status predicted <VP2>
Query Match 11.1%; Score 444.5; DB 1; Length 727;
Best Local Similarity 23.7%; Pred. No. 3.7e-21;
Matches 187; Conservative 117; Mismatches 295; Indels 189; Gaps 35;
Qy 43 RGLVLPGYKYLGPFGNLDKGEVPVNAADAAALSHDKAYDQCLKAGNPYLRYNHADAEFOE 102
Db 10 RGLVPPGYKYLGPFGNLDQGEPTNPSDAAAKEHDEAYAAVLRSGKNPYLYFSPADQRFID 69
Qy 103 RLQEDTSFGGNLGRAVFOAKKRVLPGLGLVEEGAKTAPGKKRPVQSPQEPDSSSGIGKT 162
Db 70 QTKDATWGGKIGHYFFRAKATAPVLTDPDHPSTSRPTK-PTKRSKPPPHIFINLAK- 127
Qy 163 GQOPAKRLNFGQTDGSESVP-----DQPLGEPPA--TPAAVGTPTMASGGGAPMADNNE 216
Db 128 -----KKKAGAGQVKRDNQAPMSDGAQVDPDGGQPAVNERATGSGNSGGGG-----GG 176
Qy 217 GADGVGNAGNWH-----CDSTWLGDVRVITSTRTWALPTYNNHLYKOI-----SS 262
Db 177 GSGGVGISTGTFFNNQTEFKFLENGWV--EITANSRLVHLNMPSESYKRVVNNMDKTA 234
Qy 263 ASTGASNDNHVFGYSTPWGYDFNRFHCHFSRQWQRLNNWGRPKRLNFKLFNIQV 322
Db 235 VKGNMALDDTHQVIVTPMSLVNANAWGVFNFGDWQLIVNTMSBLHLVFSFQEIFNVVLK 294
Qy 323 EVT---TNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCL-----PPFPA----- 368
```

295 TVSESATQPTKVVYNDLTASLVALDSNNMTPTPAARSETLGFYFWKPTIPTPWRY 354  
369 ---DVFMIPQVGYLTLLNGSAQVRS-----SFYCLE-YFPQMLRTGNP-TFS 413  
355 FQWDRTLIPSH-----TGTSGTPTNIYHGTPDDVQFYTIENSVPVHLLRTGDEFATG 408  
414 YTFEEVPHSSYAHQSLSLDRMN--PLIDQVLYLNRNTQSGSAQNKDLFFSRGSPAGM 471  
409 FFPDCKP--CRLTHWTQNRALGLPP-----FLNSLPQSEBGTNFGDVGQDQKRGV 459  
472 S-----VQPKN-WLPGCYRQORVSK-----TKTDNNNSNFTWT 504  
460 TQMGNTDIYIATEITMRPAEVCYSAPYISFEASTQPKPIAAGRGGAQTDENQAA---D 516  
505 GASKY---NLNGRESIINPGT---AMASHKDDDEKFFPMGVMIFGKESAGASNTALD- 556  
517 GDRYAFGRHGQKTTTGTETPERFTYIAHQT-----GRYPAGDWIQNINF 563  
557 NVMITDEEIKATNPVATERFGTVAVNPFQSSSTDPATGDVHAMGALPGMWQDRDVLQ 616  
564 NLPVTNDVLLPTDPIG---GKTGINY--TNIENTYGLTALNNVP-----PVYPNG 610  
617 PIWAKIPIHTDGHFHPSPMLGFGFLK-----NPPQILIKNTVPVPA--PPA 660  
611 QIWDKEFDTD-----LKPRLHVNAFPVCQNCPCGQLFVKVAPNLITNEYDPDA 657  
661 EPSATKPAFTIQTSTQGVSVIEIWELOKESKRWNEVQYTSYAKSANVDFTVDNGL 720  
658 SANMSR---IVTYSDFWKGKLVFKAKLRASHTWNPQQMSIN-----VDNQFN 703  
721 YTEPRPIG 728  
704 YL-PNNIG 710

RESULT 8  
VCPVME  
coat protein VP1 - mink enteritis virus (strain Abashiri)  
N;Contains: coat protein VP2  
C;Species: mink enteritis virus, MEV  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 12-Apr-1996  
C;Accession: B38350  
R;Kariatsumari, T.; Horiuchi, M.; Hama, E.; Yaguchi, K.; Ishiguro, N.; Goto, H.; Shinag  
J. Gen. Virol. 72, 867-875, 1991  
A;Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the  
A;Reference number: A38350; MUID:91202123; PMID:2016597  
A;Accession: B38350  
A;Molecule type: DNA  
A;Residues: 1-722 <KAR>  
A;Cross-references: GB:D00765  
C;Superfamily: parvovirus coat protein  
C;Keywords: coat protein  
P;119-722/Product: coat protein VP2 #status predicted <VP2>

Query Match 11.0%; Score 439.5; DB 1; Length 722;  
Best Local Similarity 23.6%; Pred. No. 7.8e-21;  
Matches 186; Conservative 117; Mismatches 295; Indels 189; Gaps 35;

44 GLVLPGYKYLGPFGNLDKGEFVNADAAALHDKAYDQQLKAGDNPVLRYNHADAEPQER 103  
6 GLVPPGKYKYLPGNSLDQGEPTNPSDAAAKHDEAYAAVLRSGKNPYLYFSPADQRFID 65

104 LQEDTSFGNLRGAVFOAKKRVLEPLGLVEBEGAKTAPGKRPVEOSQPEPDSGGIGKTG 163  
66 TKDATDGGKLGHYFFRAKKAIAPVLTDTDPNFTSRTPTK-PTKRSKPPPHIFINLAK-- 122

164 QOPAKKRLNFGQTGDSSEVP-----DPQPLGEPPA--TPAAVGTPTWASGGGAPMADNNEG 217  
123 ----KKKAGAGQVKRDNLAIPMSDGAQVQPDGQPAVRNERATGSGNGSGGG-----GGG 172

218 ADGVGNASGNWH-----CDSTWLGDRVLTSTRTWALTFTYNNHLYKQI-----SSA 263  
173 SGGVGISTGTFTNNQTEPKFLENGWV--EITANSRLVHLNMPSENYKRVVANNMDKTAV 230

264 STGASNDNHYFGYSTWGYDFENRFCHPSPRDQBLNNWNGFRPKRLNFKLFNIOVKE 323  
231 KGNMALLDDTHVQIVTTPWLSVDANAWGFWNPGWQJIVNTMTSELHLVSPQEIEFNVVLKT 290  
324 VT---TNDGVTTIANNLTSTVQVFSDESYQLPVVLGSAHQCL-----PPFPA----- 368  
291 VSSATQPTKVVYNDLTASLVALDSNNMTPTPAARSETLGFYFWKPTIPTPWRY 350  
369 ---DVFMIPQVGYLTLLNGSAQVRS-----SFYCLE-YFPQMLRTGNP-TFSY 414  
351 QMDRTLIPSH-----TGTSGTPTNIYHGTPDDVQFYTIENSVPVHLLRTGDEFATG 404  
415 TFEVPHSSYAHQSLSLDRMN--PLIDQVLYLNRNTQSGSAQNKDLFFSRGSPAGMS 472  
405 FPDCKP--CRLTHWTQNRALGLPP-----FLNSLPQSEBGTNFGDVGQDQKRGV 455  
473 -----VQPKN-WLPGCYRQORVSK-----TKTDNNNSNFTWTG 505  
456 QMGNTDIYIATEITMRPAEVCYSAPYISFEASTQPKPIAAGRGGAQTDENQAA---DG 512  
506 ASKY---NLNGRESIINPGT---AMASHKDDDEKFFPMGVMIFGKESAGASNTALD-N 557  
513 DRYAFGRHGQKTTTGTETPERFTYIAHQT-----GRYPAGDWIQNINF 559  
558 VMITDEEIKATNPVATERFGTVAVNPFQSSSTDPATGDVHAMGALPGMWQDRDVLQ 617  
560 LPTNDVLLPTDPIG---GKTGINY--TNIENTYGLTALNNVP-----PVYPNG 606  
618 IWAKIPIHTDGHFHPSPMLGFGFLK-----NPPQILIKNTVPVPA--PPAE 661  
607 IWDKEFDTD-----LKPRLHVNAFPVCQNCPCGQLFVKVAPNLITNEYDPDAS 653  
662 PSATKPAFTIQTSTQGVSVIEIWELOKESKRWNEVQYTSYAKSANVDFTVDNGLY 721  
654 ANMSR---IVTYSDFWKGKLVFKAKLRASHTWNPQQMSIN-----VDNQFN 699  
722 TEPRPIG 728  
700 L-PNNIG 705

RESULT 9  
VCPVFP  
coat protein VP1 - feline panleukopenia virus (strain 193)  
N;Contains: coat protein VP2  
C;Species: feline panleukopenia virus, FPLV  
C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 12-Apr-1996  
C;Accession: B36608  
R;Martyn, J.C.; Davidson, B.E.; Studdert, M.J.  
J. Gen. Virol. 71, 2747-2753, 1990  
A;Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvo  
A;Reference number: A36608; MUID:91073139; PMID:2174965  
A;Accession: B36608  
A;Molecule type: DNA  
A;Residues: 1-727 <MAR>  
A;Cross-references: GB:X55115  
C;Superfamily: parvovirus coat protein  
C;Keywords: coat protein  
P;144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 10.8%; Score 431.5; DB 1; Length 727;  
Best Local Similarity 23.5%; Pred. No. 2.6e-20;  
Matches 185; Conservative 117; Mismatches 297; Indels 189; Gaps 35;

43 RGLVPGYKYLGPFGNLDKGEFVNADAAALHDKAYDQQLKAGDNPVLRYNHADAEPQER 102  
10 RGLVPPGKYKYLPGNSLDQGEPTNPSDAAAKHDEAYAAVLRSGKNPYLYFSPADQRFID 69

103 RQEDTSFGNLRGAVFOAKKRVLEPLGLVEBEGAKTAPGKRPVEOSQPEPDSGGIGKT 162  
70 QTKDAKDWGGKLGHYFFRAKKAIAPVLTDTDPNFTSRTPTK-PTKRSKPPPHIFINLAK- 127

QY 163 GQPAKRLNFGQTDGSESV-----DPQLGEPPA--TPAAVGPPTTMSGGAPMADNNE 216  
DB 128 -----KKAGAGQVQRDNLAPMSDGAQVDPGQAPVRNERATGSGNGSGGG-----GG 176  
QY 217 GADGVGNASGNWH-----CDSTWLGDRVITSTRTWALPTYNHLYKQI-----SS 262  
DB 177 GSGGVGISTGTNNQTEPKFLENGWV--BITANSRLVHLNPNSENYKRVVNNMDKTA 234  
QY 263 ASTGASNDHYYGYSTPMGYFDFNPHCHFSPRDQRLNINNNWGRPRKRLNPKLNIQVK 322  
DB 235 VKGNMALDDIRHQIVTPMVLVDANAWGVWVFNPDQWOLIIVNTMSELHLVSFEQIEINVVLK 294  
QY 323 EYT--TNDGVTTIANNTSTVQVSDSEYQIYVVLGSAHQGL-----PPFA----- 368  
DB 295 TVSESATQPTKVYNDLTASLWALDSNNTPFTPAAMRSSTLGPYKPKPIPTPWRY 354  
QY 369 ---DVFMIPQYGLTLNNGSQAVGRS-----SFYCLE-YFPQOMLRTGNPF-TFS 413  
DB 355 FQWDRTLIFSH-----TGTSGTPTNVHGTDPDVQFYTIENSVPVHLRTGDEFATGT 408  
QY 414 YTFEEVPHSHSAHSQSLDRLMN--PLIDQYLYLNRTQNGSGSAQNKDLLFSRGSAPGM 471  
DB 409 FPFDCPK--CRLTHWTQTNKRALGLPP-----FLNSLPQSEGATNFGDIGVQODKRRGV 459  
QY 472 S-----VQPKN-WLPGPCYRQORVSK-----TKTDNNNSNFTWT 504  
DB 460 TQNGTNDITYTEATIMKPAEYGVSAFYSPFESTQGFKTPIAAGRGGAQTENQAA---D 516  
QY 505 GASKY---NLNGRESIINFGT---AMASHKDEKDFPMPGSMVIFGKESAGASNTALD- 556  
DB 517 GDRYAFGRHQKQKTTTGTGETPERFTYIAHQDT-----GRYPEGDWQINIF 563  
QY 557 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHANGALPGVWQDRDVIYLOG 616  
DB 564 NLPVTDNDVLLPTDPIG---GKTGINY--TNIFTYGLTALNNVP-----PYFNG 610  
QY 617 PIWAKIPHDTDGHFSPMLGGFLK-----NPPQILIKNTVPVAN---PPA 660  
DB 611 QIWDKEPDT-----LKRLHVNAPFVQNNCPQLFKVYAPNLITNEYDPA 657  
QY 661 BPSATKASFITQYSTQGVSVIEWELQKNSKRMNPEVQYTSNYAKSANVDFTVNNGL 720  
DB 658 SANMSR-----IVTSDFMWKGKLVKAKLRASHATWNPQQMSIN-----VDNQFN 703  
QY 721 YTEPRPIG 728  
DB 704 YV-PNNIG 710

RESULT 10  
VCPVPP

coat protein VP1 - porcine parvovirus

N:Contains: coat protein VP2

C:Species: porcine parvovirus

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 12-Apr-1996

C:Accession: B33302

R:Ranz, A.I.; Manclaus, J.J.; Diaz-Aroca, E.; Casal, J.I.

J. Gen. Virol. 70, 2541-2553, 1989

A:Title: Porcine parvovirus: DNA sequence and genome organization.

A:Reference number: A33302; MUID:90010964; PMID:2794971

A:Accession: B33302

A:Molecule type: DNA

A:Residues: 1-723 &lt;RAN&gt;

A:Cross-references: EMBL:D00623

C:Genetics: 10/1

A:Superfamily: parvovirus coat protein

C:Keywords: coat protein

F:145-723/Product: coat protein VP2 #status predicted &lt;VP2&gt;

## Query Match

Best Local Similarity 10.7%; Score 426; DB 1; Length 723;

Matches 190; Conservative 109; Mismatches 308; Indels 156; Gaps 36;

QY 53 LGPENGLDKGPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSEGG 112  
DB 13 LGPNSLDQGEPTNPSDAAAKEHDEAYDKYIKSKNPTFYPSAADEKFIKETEHAQYGG 72  
QY 113 NLGRAVFOAKKRVLEPLGLVEEGAKTAPKKRPVQEQPQEPDSSSGIGK-----T 162  
DB 73 KIGHYFPAKGAFAFKL---SETDSPTTSQOPEVRSRPHKPGPKGKRPAPRHIFNL 129  
QY 163 GQPAKRLNFGQTDGSESVDPDQPLGEPPATP-AAVPTTMSGGGAPMADNNEGADGV 221  
DB 130 AKKAKGTSNTNSMSGENVEQHNPIN--AATELSATGNSGGGGGG-----GGRGAGV 183  
QY 222 GNASGNWH--CDSTWLGQ---RVITSTRTWALPTYNHLYKQI-----SSASTGAS-ND 270  
DB 184 GVSTGSFNNQTEFYQYLGELVRITAHASRLIHLNPNHEHYTKRIHVLNSBSGSAGMQVD 243  
QY 271 NHYFGYSTPMGYFDFNPHCHFSPRDQRLNINNNWGRPRKRLNPKLNIQKVEVT---TN 327  
DB 244 DAHTQMTVPMSLIDANAWGVWVFNPDQWOLIINNTEINLVSEQAI FNVVLKTTITESATS 303  
QY 328 DGVTTIANNTSTVQVSDSEYQIYVVLGSAHQGLPPFPADVFMIPQYGL----- 379  
DB 304 PPTKIYNDLTASLWALDNTNLTPTPAAPRSETLGFYPMWLPKTPQYRYLSCLRNLN 363  
QY 380 --TLNNGSOAVGRSS-----FYCLE-YFPQOMLRTGNPF-TFSYTFEEVPHSSYA 426  
DB 364 PPTYTGSGQPNRLNTRLNRLHSDINFYTIENAVPIHLRTGDEFSTGIYHPTKPL--KLT 421  
QY 427 HSQSLDR-----LMNPLI--DQYLYLNRTQNGSGSAQNKDLLFSRGSAPGMSVOPK 476  
DB 422 HSWQTNRSGLPKLLTEPTTEGQHPGTLPAATRKYHQTIINNSYTEAT---AIRP- 476  
QY 477 NWLPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMASHKDEKDF 534  
DB 477 -----AQGVNTPYMNFYSNGGPF-----LTPIVPTADQYNDDBSN- 514  
QY 535 FPMGSMVIFGKESAGASNTALD---NVMITDEEIK--ATNP-----VATERFGTVAV 582  
DB 515 -----GATFTMDYOHGHLTTSQBELRYTFNPQSKCGRAPKQFNOQAPL 560  
QY 583 NFQSSS-----TDPATG--DVHANGALP-----GMVWQDRDVIYLOGFIWAKIPHDT--G 627  
DB 561 NLENTNNGTLLPSDPIGKSNMFMWTLNTYGLTALNNTAPVPNGQIWDKELDTLKP 620  
QY 628 HFHPSPLMGFGFLK-NPPQILIKNTVPVANPPAEFSA-TKFASFITQYSTQGVSVIEW 685  
DB 621 RLH---VTAPFVCKNNPFGQLFVKIAP---NLTDQFNADSPQQPRIITYSNFMWKGTLTF 674  
QY 686 ELQKNSKRMNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIG 728  
DB 675 TAKRSSNNWNPQQHTT-----TAENIGKYI-PTNIG 706

RESULT 11  
VCPVIM

coat protein VP1 - minute virus of mice (strain MVM1)

C:Species: minute virus of mice, murine parvovirus

C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 28-Jul-2000

C:Accession: B23008; B29510

R:Sahli, R.; McMaster, G.K.; Hirt, B.

Nucleic Acids Res. 13, 3617-3633, 1985

A:Title: DNA sequence comparison between two tissue-specific variants of the autonomous 1

A:Reference number: A23008; MUID:85242059; PMID:3855242

A:Accession: B23008

A:Molecule type: DNA

A:Residues: 1-718 &lt;SAH&gt;

A:Cross-references: EMBL:X02481; NID:G60918; PIDN:CAB46507.1; PID:G5419928

J. Virol. 57, 656-669, 1986

A:Title: DNA sequence of the lymphotropic variant of minute virus of mice, MVM(1), and c

A:Reference number: A29510; MUID:86115415; PMID:3502703

A:Accession: B29510

A:Molecule type: DNA  
A:Residues: 1-143,'A',145-718 <AST>  
A:Cross-references: EMBL:M12032  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein

Query Match 10.6%; Score 424.5; DB 1; Length 718;  
Best Local Similarity 23.7%; Pred. No. 7.4e-20;  
Matches 184; Conservative 104; Mismatches 298; Indels 189; Gaps 31;

```
QY 45 LVLPGYKLPFGNLDKGEVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAERL 104
Db 1 MWPGGYKLPFGNLDKGEVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAERL 104
QY 105 QEDTSFGNLRGAVFOAKRVLEPLGLVEEGAKTAPKCRPVESQSPQEDSSSGIGKTG 163
Db 61 KOAKDGGKGVGHYFFRTKRAFAKPLATDSE-----PG-----TSGVSRACK 101
QY 164 -----QOPAKRLNFG---QTGDSSEVPDPQPLG-----EPPATPAAVGPTTM 203
Db 102 RTRPPAYIFINQARAKKCLTSSAAQSSQTSMDTSGPDGNGVHSAARVERAADPG-- 159
QY 204 ASGGGAPMADNEGAGVGNASGNHCHDS--TWLGD---RVITSTRTWTALPTYNHLYK 258
Db 160 GSGGG-----GSGGGGVSTGSDYDNQTHRYLFGDWVEITATRLVHLNMPKSENYC 213
QY 259 QI-----SSASTGASNDNHFGYSTPMGYFDNFRHCFHSPRDQRLINNNWGRPKR 311
Db 214 RIRVHTTDTSVKGNMAKDAAHEQIWTPLSLVDANAWGWLQPSDQYICNTMSQLNLVS 273
QY 312 LNFKLNIQKVEVTND-----GVTTTANNLTSTVQVPSDSEYQLPVYLSAHQGLPPPP 367
Db 274 LQOEIENVVLKVTQEDSGQAIKIYNNDLTACMVAVDSNNILPYTPAANSMETLGFYP 333
QY 368 ADVEMIPQYGY-----LTLNN-----GSQAVGRSSPYCLEYFVS--QMLRTG 407
Db 334 WKPTIASPRYFCVDRDLSVTYENQEGTIEHNWGTPKGMSQFTTIENTQOITLRTG 393
QY 408 NNF-TFSYTFEEVPHSSVAHQSRLDMN--PLIDQYLYLNRQNSQSAQNKDLLFS 464
Db 394 DEPATGYTFDTPV--KLTHWTQTNKQLGQPLLSTF-----PEADTDAGT-----LTA 441
QY 465 RGSFAGMSVQPKWL-----PGCYRQORVSKTKTDNNNSN 500
Db 442 QGSRHGATQEMVNWSEAIRPAQVGFQCPHNDFEASRAGP--FAAPKVPADVTCQVDRE 500
QY 501 FTWTGASKYNLNGRESIINPGTAMASHKDDKDF-----FPMMSGVMIFGKESAGAS 551
Db 501 ANGSVTVSYGKHGEGNWAHGAHPAPERYTWDETNGSGRDRDGTQFIOQAPLVPPPLNGIL 560
QY 552 NTALDNVMTDDEEIKATNPVATERFGTVAVNFQS--SSDTPATGDVHAMGALPGMVWD 609
Db 561 TNA-----NPIGKN---DIHFSNVFNSVGLPTAFSH-----P 590
QY 610 RDVYLOGPIWAK---IPHTDGHFHPSPMGGLKNPPQIILKNTP--VPANPPAEFSA 664
Db 591 SPVYPOGQIWDKELDLHKLPRHLITAFV--CKNNAPQQLVRLGNLTQDYPNGATL 647
QY 665 TKFASITQVSTQGVSEIWELEQENSKRNVEVOYTSNYAKSANVDFTVDNNG 719
Db 648 SRIVTGTFFWKGKLTWRKILRA-----NTWNVPYQ-----VSVEDNG 686
```

RESULT 12  
VCPVCP

coat protein VP1 - canine parvovirus (strain N)  
N:Contains: coat protein VP2  
C:Species: canine parvovirus, CPV  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 12-Apr-1996  
C:Accession: B29962  
J:Reed, A.P.; Jones, E.V.; Miller, T.J.  
J. Virol. 62, 266-276, 1988  
A:Title: Nucleotide sequence and genome organization of canine parvovirus.

A:Reference number: A29962; MUID:88062392; PMID:2824850  
A:Accession: B29962  
A:Molecule type: DNA  
A:Residues: 1-748 <REE>  
A:Cross-references: EMBL:M19296  
C:Genetics:  
C:introns: 26/3  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F:165-748/Product: coat protein VP2 #status predicted <VP2>

Query Match 10.6%; Score 421.5; DB 1; Length 748;  
Best Local Similarity 23.0%; Pred. No. 1.2e-19;  
Matches 184; Conservative 114; Mismatches 287; Indels 215; Gaps 34;

```
QY 44 GLVLPGYKLPFGNLDKGEVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAERL 103
Db 32 GLVPPGYKLPFGNLDKGEVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAERL 103
QY 104 QEDTSFGNLRGAVFOAKRVLEPLGLVEEGAKTAPKCRPVESQSPQEDSSSGIGKTG 163
Db 92 TKDADWGGKLGHYFFRAKKAIPVLTDPHPSTSRPTK-PTKRSKPPHIFINLAK-- 148
QY 164 QOPAKRLNFGQTSSEVSP-----DQPLGEPBA--TPAAVQPTTMSAGGAPMADNNEG 217
Db 149 -----KKKAGAGQVGRDNLAPMSDGAVPDGGQPAVRNERATGSGNGSGGG--GGG 198
QY 218 ADGVGNASGNWH-----CDSWTGLGDRVITTTSTRTWTALPTYNHLYKQI-----SSA 263
Db 199 SGVGIISGTGTTNQTPEKPLENGWV--EITANSSRLVHLNMPESERYRVVNNMDKTAV 256
QY 264 STGASNDNHFGYSTPMGYFDNFRHCFHSPRDQRLINNNWGRPKRLNFKLNIQVKE 323
Db 257 NGNMALDDIHAQIVTPWLSLVDANAWGVNFGDWQLIVNTWSELHLSVFEQEI FNVVLKT 316
QY 324 VT---TNDGVTTIANNLTSTVQVPSDSEYQLPVYLSAHQGL-----PPPPA----- 368
Db 317 VSESATQPTTKYNNNDLTASLWALDLSNNTMPTTPAAMRSETLGFYPMKPTTPTWRYF 376
QY 369 --DVFMIPOGYLYTLNNGSQAVGRS-----SFYCLE-YFPSQMLRTGNPF-TPSY 414
Db 377 QWDRTLIFSH-----TGTSGTPTNIIYHGTDPDDVQVFTIENSVPVHLLRTGDEATGTF 430
QY 415 TFEVFPFHSSVAHQSRLDMN--PLIDQYLYLNRQNSQSAQNKOLLFSRSGSPAGMS 472
Db 431 PFDCPK--CRLTHWTQTNKALGLP-----FLNSLPQSEGATNFGDI-----GV- 472
QY 473 VQPKWLPGCYRQORVSKTKTDNNNSN-----TWTCASKYNLNGRESIINP-GT 522
Db 473 -----QODKRGVQMGNTNYITEATIMRPAEVGYSAPYYSFEASTQGPFKT 519
QY 523 AMASHK-----DDEDKFPMSGVMIFGKESAGASNTALD----- 556
Db 520 PIAAGRGGAQTYENQAADGPRY-----AFGRHQGKTTTGTPTERTVIAHQDQTR 572
QY 557 -----NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPG 604
Db 573 YPEGDWIQINFNLPVTDNVLPLTDPIG---GKTGINY--TNIFNTVGLPTALNNVP- 625
QY 605 MWQDRDVLQPIWAKIPIHTDGHFHPSPMGGLKNPPQIILKNTP--VPANPPAEFSA 651
Db 626 -----PVYPNGQIWDKEFDTD-----LKPRLHVNAPFVCCNNCPGQLFVKV 666
QY 652 TPVPAN---PPAEFSATKFAFGITQYSTQGVSEIWELEQENSKRNVEVOYTSNYAKS 708
Db 667 ABNLTVEDVDPASANMSR-----IVTYSDFWKGKLVFKAKLRASHTWNPDIQMSIN----- 718
QY 709 ANVDFTVDNNGLYTEPRPIG 728
Db 719 -----VDNQFNIV-PSNIG 731
```

RESULT 13

## VCPVCD

coat protein VP1 - canine parvovirus (strain CPV-d)  
N:Contains: coat protein VP2  
C:Species: canine parvovirus, CPV  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999  
C:Accession: A31163  
R:Parish, C.R.; Aquadro, C.F.; Carmichael, L.E.  
Virology 166, 293-307, 1988  
A:Title: Canine host range and a specific epitope map along with variant sequences in the  
A:Reference number: A31163; MUID:89020796; PMID:3176341  
A:Accession: A31163  
A:Molecule type: DNA  
A:Residues: 1-737 <PAR>  
A:Cross-references: EMBL:N23255; NID:g333467; PIDN:AAA47158.1; PID:g333468  
C:Genetics:  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F:584-737/Product: coat protein VP2 #status predicted <VP2>

Query Match 10.2%; Score 407.5; DB 1; Length 737;  
Best Local Similarity 22.8%; Pred. No. 9.9e-19;  
Matches 181; Conservative 112; Mismatches 301; Indels 201; Gaps 32;  
QY 42 GRGLVLPQYKYLGPFGNLDKGEVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAEFQ 101  
DB 19 GEGKDLITYKYLGPFGNLDKGEVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAEFQ 78  
QY 102 ERLQEDTSGGNLGRAVFOAKKRVLEPLGLVEEAGTAPGKRPVQSQPQSDSSGIGK 161  
DB 79 DQTKDAKDWGKLGHYFFRAKAIAPLVTDTPHSTSRPTK-PYKSKPPPHIINLAK 137  
QY 162 TGOQPAKRLNFGQTDSSSVP-----DQPLGEPPA--TPAAVGPTTMASGGAPMAADNN 215  
DB 138 -----KKAGAGQVQRDLAPMSDGAQVQDQGPAPVRNERATGSGNGSGGG-----G 185  
QY 216 EGADGVGNAGNWH-----CDSTWLGDRVITTSRTWALPYNHLYKQI-----S 261  
DB 186 GSGGVGISTGTNNTEPKFLENGWV--EITANSSRLVHLAMPSENRVRRVNNMDKT 243  
QY 262 SASTGASNDNHVFGYTPMGYDFNRFCHFSRDMQRLNNWGPRLNFKLFIQV 321  
DB 244 AVNGNMAADDIAHQIVTPSLVDANAGWVFPDQWLIVNTMSELHLVSPQEIFNVVL 303  
QY 322 KEVT-----TNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQGL-----PPFPA----- 368  
DB 304 KTVSESATOPPTKVYNNDLTASLMVALDSNNTMPTPAAMRSETLGFYPKWPTIPTPMRY 363  
QY 369 ----DVFMIPQYGYLTLLNGSQAVGRS-----SFYCLR-YPPSQMLRTGNF-TF 412  
DB 364 YFQWDRTLPSH-----TGTSGTPTNIYHGTDPPDQVFTTENSVPVHLLRTGDEFATG 417  
QY 413 SYTFEEVPFHSSYAHQSQDLRLMN--PLIDQVLYLNTQNSGSAQNKDLFLSRGSPAG 470  
DB 418 TPFDFCKP--CRLTHTWQTNKALGLPP-----FLNSLPQSEGATNFGDI-----G 460  
QY 471 MSVQPKMLPCPCYRQORVSKTKTDNNNF-----TWGASKYNLNGRESIINP- 520  
DB 461 V-----QODKRGVTOGMNTNYITEATIMRFAEVGYSAFYSPASTQGP 506  
QY 521 ----GTAMASHKDEDEKFFPMGSMVIFGKESAGASNTALD----- 556  
DB 507 KTIPIAGRGAQTDENQAADGNPRYAFGRHQGKTTTGTETPERFYIAHQDTGRYPEGD 566  
QY 557 -----NVMITDEEIKATNPATERFGTVAVNFQSSSTDPATGDVHANGALPGMVWQD 609  
DB 567 WIQINFNLPVTDNVLPTDPIG-----GKTGINY--TNIFTYGLPTALNNVP----- 614  
QY 610 RDVYLOQPIWAKIPHTDCHPHPLMGGFGCLK-----NPPQILIKNTVPVA 656  
DB 615 -PVYPNGQIWDKEFDTD-----LKPRLHVNAPFVQNNCPQGLFVKVAPNLT 660  
QY 657 N---PPAEFSATKFAFITQYSTGQSVSEIWELOKENSCKRMNPEVQYTSNYSKASNVDF 713

DB 661 NEYDPDASANMSR-----IVTYSDFWMMKGLVFAKULRASHTWNPQQMSIN----- 707  
QY 714 TVDNNGLYTBPRPIG 728  
DB 708 -VDNQFNIV-PSNIG 720  
RESULT 14  
VCPV2  
coat protein VP1 - parvovirus H1  
C:Species: parvovirus H1  
A:Note: host Homo sapiens (man)  
C:Date: 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change 08-Apr-1994  
C:Accession: A03699  
R:Rhode III, S.L.; Paradiso, P.R.  
J. Virol. 45, 173-184, 1983  
A:Title: Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid  
A:Reference number: A03695; MUID:83112183; PMID:6823009  
A:Accession: A03699  
A:Molecule type: DNA  
A:Residues: 1-722 <RHO>  
A:Cross-references: EMBL:X01457; EMBL:J02198  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
Query Match 10.1%; Score 404; DB 1; Length 722;  
Best Local Similarity 23.3%; Pred. No. 1.6e-18;  
Matches 175; Conservative 108; Mismatches 290; Indels 178; Gaps 30;  
QY 57 NGLDKGEVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAEFQERLOEDTSPGNGLR 116  
DB 12 NSLQGEPTNPSDAAKHEDEAYDYIKSGKPYLYFSPADQRFIDQTKADKDWGKVG 71  
QY 117 AVFOAKKRVLEPLGLVEEAGTAPGKRPVQSQPQSDSSGIGTKGQ-----PA----- 167  
DB 72 YFFRTKRAFAPKLTDS-----EG-----TSGVSRPGKRTKPPAHIFVQ 112  
QY 168 -----KRLNFCQTDSSSVPPQPLGEPPTAPAAVGPTTMASGGAPMAADNNEGADGVN 223  
DB 113 ARAKKGRASLAQAQRTLTMSDGTETNPDGTIANARVERSADGGG--SGGGSGGGGIGV 171  
QY 224 ASGNWHCDSTW--LGDRTVITTSRTWAL-----PTYNHLYKQISASTGA 267  
DB 172 STGYDNTQTTKFLGDGHWETATASRLHLGMPSENYCRVTVHNNQTTGHTKVKGNM 231  
QY 268 SNDNHYFGYSTPWGYDFNRFCHFSRDMQRLNNWGPRLNFKLFIQVKEVT-- 326  
DB 232 AYDTHQIWI--TPWSLVDANAGWVFPDQWQFIQNSMESLNLDSLSQELFNVVVKTVEQ 290  
QY 327 ----NDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQGL-----PPFPAD-----VPMI 373  
DB 291 QGAGQDALIKVYNNDLTACMMVALDSNNILPTTPAAQTSETLGFYPKWKTAPAPYRYFFM 350  
QY 374 PQYGYLTLLNGSQ-----AVG-----RSSFYCLR-YPPSQMLRTGNFTF--SYTFEEVP 420  
DB 351 PQLSVTSSNAGEQTQITDTIGEPAQNSQFTIENTLPITLLRTGDEFTTGYIFNDP 410  
QY 421 FHSSYA-----HSQSLDRLMN-PLIDQVLYLNTQNSGSAQNKDLFL----- 463  
DB 411 LKLTHTWQTNRLACLQITDLPISDTATASLTANGDRFGSTQTONVNVYVTEALRTPAQ 470  
QY 464 -----SRGSPAGMSVQPKMLPGCYRQORVSKTKTDNNNF--TWGASKYNL 511  
DB 471 IGFMPHONFENRGGPFKVPVPP-----LDITAGEDHDAN-----GAIRFNY 513  
QY 512 NGR--ESIINPGTA-----MASHKDEDEKFFPMGSMVIFGKESAGASNTALDNNV 559  
DB 514 GHQGEDWAKQGAAPERITWDAIDSAAGBDTARCFV-----QSAPISI PPNQOI 563  
QY 560 ITDEEIEKA--TNPATERFGTVAVNFQSSSTDPATGDVHANGALPGMVWQDRDVLQGI 618  
DB 564 LQREDAIAGRITNMHTYTNFNSYGLPSAFPHPDP-----IYPNGOI 603

[illegible]

**RESULT 15**

VCPVCN  
 coat protein VP1 - canine parvovirus (strain 780929)  
 N:Contains: coat protein VP2  
 C:Species: canine parvovirus, CPV  
 C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 12-Apr-1996  
 C:Accession: A03702  
 R:Rhode III, S.L.  
 J. Virol. 54, 630-633, 1985  
 A:Title: Nucleotide sequence of the coat protein gene of canine parvovirus.  
 A:Reference number: A03702; MUID:85185696; PMID:3989914  
 A:Accession: A03702  
 A:Molecule type: DNA  
 A:Residues: 1-722 <RHO>  
 A:Cross-references: EMBL:M10989  
 C:Genetics:  
 A:Introns: 11/1  
 C:Superfamily: parvovirus coat protein  
 C:Keywords: coat protein  
 F:139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match 9.7%; Score 385.5; DB 1; Length 722;  
Best Local Similarity 22.8%; Pred. No. 2.6e-17;  
Matches 178; Conservative 118; Mismatches 297; Indels 189; Gaps 35;

Qy	49	GKYLGPFGNGLD	KGEPVNAADAAL	EHDKAYDQOL	KAGDNPYLRYNHADAE	FOERLOEDT	108
Db	11	GKYLGPFGKSL	DOGETPNPSDAA	AKHEDEAVAA	YLRGKNIPYLY	FSPADQRFID	109
Qy	109	SFGCNIGRAV	FOAKKRVLEPL	GLVBEAGAKTAP	KKKRPVEQSPQ	BSDSSGIGTK	168
Db	71	DWGGKIGHYF	FRAKKAIAPVL	TDTDPHSTSR	PTK-PTKSKPP	PHHFINLAK	169
Qy	169	KRLNFQGTG	DESVP-----	DPOPLGEPPA	---TPAAVGFT	WASGGGAPMAD	222
Db	124	KKAGAGQVR	KDNLAPMSDGA	VPDGGQPAV	NERATGSGNG	SGGGG-----	223
Qy	223	NASGNWH-----	CDSTWLGDR	VTTSTRTWAL	PTNNHLYKQI	-----SSASTGAS	268
Db	178	ISTGTENNQ	TEPKFLENGWV	---EITANSRL	VHLHMPESK	ORVVVNNMDK	235
Qy	269	NDNHFYGY	SPMGYDFNRF	CHCFSPDWO	RBLNNMGFAP	KRLNFKLFI	325
Db	236	LDDIHAQI	VTVPMSLV	DANAWDV	WFPNGDW	QLIVNTMSL	495
Qy	326	TNDGVTTI	ANNLTSTVO	VSFDSSEYQ	LYVLGSANQ	GGCI-----	371
Db	296	TOPPTKY	VNNDLUTAL	UVALDSNN	TMPTPAAR	SETLGPYK	355
Qy	372	MIPOGYLT	LNNGSQAVGRS	-----SFYCLE	-YFPQSOML	TGNNE-TFSY	419
Db	356	LIPSH-----	TGTSPTNII	VHGTD	PDQVFIYI	ENSPVPHLL	409
Qy	420	PFHSSVAH	SQSJDLRL	MNPLIDQ	LYLNRNTQ	S-----GSAQNK	460
Db	410	P-CRLLTH	TWQTNRALG	--LPPFLN	SLPSQSE	GATNFIDIG	465
Qy	461	LLFSRGS	PAGMSVQ-----	PKNWLP	GPCVQRQ	RVSKTKTDNN	509
Db	466	ATIMRPA	EVGYSA	PYSPSEAST	QGFKTLPT	IAAGR-----	517
Qy	510	--NLNGRES	IINPGT-----	AMASHK	DDDBK	DFPMSGV	562
Db							

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 15:59:46 ; Search time 12.6014 Seconds  
(without alignments)  
2746.653 Million cell updates/sec

Title: US-09-807-802a-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWNLENSIGR.....NNGLYTEPRPICTRYLTRPL 736

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2177.5	54.6	504	COA3_AAV2	P03135 adeno-assoc
2	798.5	20.0	673	COAT_PAVBO	P07297 bovine parv
3	600.5	15.1	781	COAT_PAVHB	P07299 human parv
4	482	12.1	729	COA1_PAVP9	P33484 porcine par
5	482	12.1	729	COA1_PAVPN	P18546 porcine par
6	480	12.0	729	COA1_PAVPK	P52501 porcine par
7	444.5	11.1	727	COAT_FPV	P04864 feline panl
8	437.5	11.0	722	COAT_MEVA	P27437 mink enteri
9	431.5	10.8	727	COAT_FPV19	P24840 feline panl
10	430.5	10.8	727	COAT_PAVCB	Q11213 canine parv
11	424.5	10.6	718	COAT_MUMIM	P07302 murine minu
12	421.5	10.6	748	COAT_PAVCN	P12930 canine parv
13	407.5	10.2	737	COAT_PAVCD	P17455 canine parv
14	404	10.1	722	COAT_PAVHH	P04863 hamster par
15	386.5	9.7	722	COAT_PAVC7	P03136 canine parv
16	377	9.5	716	COAT_MUMIV	P03137 murine minu
17	234	5.9	587	COAT_PAVL3	P06310 parvovirus
18	207.5	5.2	809	COAT_DSDNV	O71155 diatraea sa
19	201	5.0	811	COAT_GMDNV	Q90125 gallitaea me
20	199.5	5.0	810	COAT_JCDNV	Q90053 junonia coe
21	195	4.9	584	COAT_PAVC2	P30129 canine parv
22	186.5	4.7	647	COAT_ADVG	P24029 aleutian mi
23	143.5	3.6	648	FXN1_MOUSE	Q61575 mus musculi
24	131.5	3.3	1446	1E18_PVKA	P33479 pseudorabie
25	126.5	3.2	1461	1E19_PVIVF	P11675 pseudorabie
26	124.5	3.1	2493	1CYAA_USTMA	P49606 ustilago ma
27	123	3.1	648	FXN1_HUMAN	O15353 homo sapien
28	123	3.1	1142	1ENAM_PIG	O97939 sus scrofa
29	121	3.0	880	1SYV_BACST	P11931 bacillus st
30	121	3.0	1113	1N1L6_YEAST	Q02630 saccharomyc
31	120	3.0	1379	1YFF9_SCHPO	O14066 schizosacch
32	119.5	3.0	1394	1HAP_HASIN	P45387 haemophilus
33	119.5	3.0	2004	1YP73_YERPE	Q8zdj2 yersinia pe

34	119	3.0	667	1	ITP2_HUMAN	P15884 homo sapien
35	117.5	2.9	1076	1	NUP1_YEAST	P20676 saccharomyc
36	116.5	2.9	642	1	ITP2_CANFA	P15881 canis famli
37	114.5	2.9	802	1	PAC_BACME	Q60136 bacillus me
38	114	2.9	1778	1	N189_SCHPO	Q9utk4 schizosacch
39	113.5	2.8	1324	1	IRS2_HUMAN	Q9y4h2 homo sapien
40	113	2.8	781	1	SP3_HUMAN	Q02447 homo sapien
41	113	2.8	837	1	HFC2_HAEIN	P45997 haemophilus
42	113	2.8	1484	1	CES2_HUMAN	Q9bxf3 homo sapien
43	112	2.8	774	1	FECA_ECOLI	P13036 escherichia
44	111.5	2.8	815	1	PYGO_DROME	Q9v9w8 drosophila
45	111.5	2.8	2944	1	CA17_HUMAN	Q02388 homo sapien

ALIGNMENTS

```
RESULT 1
COA3_AAV2
AC COA3_AAV2 STANDARD; PRT; 504 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 01-NOV-1991 (Rel. 20, Last annotation update)
DE Probable coat protein 3.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83164299; PubMed=6300419;
RA Srivastava A., Lueby E.W., Berns K.I.;
RT "Nucleotide sequence and organization of the adeno-associated virus 2
   genome.";
RL J. Virol. 45:555-564 (1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01901; AAA42376.1; -
DR PIR; A03698; VCPV3A.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein.
SQ SEQUENCE 504 AA; 56366 MW; 758999B017052B6B2 CRC64;
```

```
Query Match 54.6%; Score 2177.5; DB 1; Length 504;
Best Local Similarity 82.5%; Pred. No. 1.3e-126;
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;

QY 203 MASGGAPMADNNEGADGVGNAGSNHWCDSWLGDVRVITTTTRTWALPTNNHLYKQISS 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MATGGAPMADNNEGADGVGNSSGNHWCDSWTWGDVRVITTTTRTWALPTNNHLYKQISS 60

QY 263 ASTGASNDNHYPGYSTPMGYDFDPRFCHFSFPRDQWRQRLNNNMGPRPKRLNFKLFIQVK 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 QS-GASNDNHYPGYSTPMGYDFDPRFCHFSFPRDQWRQRLNNNMGPRPKRLNFKLFIQVK 119

QY 323 EYTTNDGVTIANNTLSTVQFSDSEYQPVYLGSAHQGCLPPFPADVPNIQYGYLTUN 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 120 EYQNDGVTIANNTLSTVQFSDSEYQPVYLGSAHQGCLPPFPADVPNIQYGYLTUN 179

QY 383 NGSQAVGRSSFCYCLYFFFSQMLRTGNNTFSTVFPEVPHSSYAHQSQSLDRMLNPLIDQY 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 180 NGSQAVGRSSFCYCLYFFFSQMLRTGNNTFSTVFPEVPHSSYAHQSQSLDRMLNPLIDQY 239

QY 443 LYYLNRTOQSGSAQNKDLFLFRSGSPAGMSVQPKWLFQPCYRQORVSKTKTDNNNSNFT 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 240 LYLISRTNTPSGTTTQSRLOFSQAGASDIRQSRNWLPGPCYRQORVSKTSADNNNSEYS 299
Qy 503 WTGASKYNLNGRESITINPTAMASHKDEDEFFPMGSGVMIRGKESAGASNTALONVMITD 562
Db 300 WTGATKYLHNGRDLNVP--AMASHKDEDEFFPMGSGVLIIFGKQSEKTNVIEKWMITD 357
Qy 563 EEEIKATNPVATERPGTVAVNFQSSSTDPATGDVHAMGALPGMVQDRDVLQGPWAKI 622
Db 358 EEEIGTNPVATEQGSVSTNLQRNQAATADVNTQGLVPGMVQDRDVLQGPWAKI 417
Qy 623 PHTDGHFPLSMGSGFGLKNBPQIILKNTVPANPPAEFSAFASITQYSTG 677
Db 418 PHTDGHFPLSMGSGFGLKNBPQIILKNTVPANPPAEFSAFASITQYSTG 472

RESULT 2
COAT_PAVBO STANDARD; PRT; 673 AA.
AC P07297; Q84374;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP2 [Contains: Coat protein VP3].
OS Bovine parvovirus (BPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10784;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=87061184; PubMed=3783814;
RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,
RA Bates R.C.;
RT "Complete nucleotide sequence and genome organization of bovine
RT parvovirus."
RL J. Virol. 60:1085-1097(1986).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M14363; AA859848.1; -
DR EMBL; M14363; AA859849.1; -
DR PIR; A26104; VCPVB5.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 673
FT CHAIN 138 673 COAT PROTEIN VP3.
FT CARBOHYD 343 343 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 446 446 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DOMAIN 163 187 GLY-RICH.
SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 20.0%; Score 798.5; DB 1; Length 673;
Best Local Similarity 28.6%; Pred. No. 9.1e-42;
Matches 214; Conservative 105; Mismatches 275; Indels 155; Gaps 21;

Qy 30 PKPKANQKQDDGRGLVPGYKLGFPNGLDKGEVPNAADAALEHDKAYDQOLKAGDNP 89
Db 3 PTNKANSKK-----GLTLFGYNLGFPSNLFAGAPVAKDAARAKHDFCYSDLLKEGKNP 57
Qy 90 YLYRNHADAERFOERLQEDTSFGNTGRVFAQKRVLEPLGLVEGAKTAPGKKRPVQS 149
Db 58 YLYFNTHDQNLDELKDDTSFGKLGARGVFQIKALAPAL-----PGTSKGGDRALKRK 111
Qy 150 PQEPSSSSGIGTKGQPAKRLNFGQTGDSVSPDQPLGEP--PATPAVGTPTWASGG 207

```

```

Db 112 LYPARSKGAKKANREPAPSTSNQONMEVSNIDIPNDEAGNQPIELATRSVVGSGVGGG 171
Qy 208 GAPMADNNEGADGVNAGSNHWCDSWLGDRVLTSTRTWALPTYNHLYKQISSASTGA 267
Db 172 -----RGSGVGSTGGTGTTFSENIIVTKTRQFIDIKNGHLYS-EVLTGSD 222
Qy 268 SNDNHFGYSTPGYDFDNRFCHFSFRDQRLNNNNWGRPKRLNPKLFNIQVKEVTN 327
Db 223 TAHRQY-AITTPWSYFNQYSSHSFSPNDQHLVNDYERFRPKAMIVRVNQLKIQIMTD 281
Qy 328 DGVTTIAN--LTSTVQVFSSEYQLPYVLSAHQGLPPPPADVFMIPOYGYLTL----- 381
Db 282 GAMGTVTYNNDLTAGMHIFCDGHRYPYVQHPDDQCMPELNSIWELPQYAYIPAPISV 341
Qy 382 -NNGSOAVGRS-----SFYCLEYPPSQMLRTNNTFSTYTFEEVPHSSVAHSQSLDRLM 435
Db 342 DNTTNTVBEHLKGVPLYMLENSDHEVLNG-----RIV 376
Qy 436 NPLIDQYLYLNRQN-----QSGSAQNKDLLEFSRGP-----AGMSVQPKWLP 480
Db 377 RIVQLWRLMRDKRKHQIHASDDVQSTGQKNLLIQRTKQPNKORFQNAALRTSNWS 436
Qy 481 GPCYRQORVSKTKDNNNSNFTWTSKYNLNGRESIINPTAMASHKDEDEKFFPMG 540
Db 437 GP-----GIARGTHNATLQTSAGALVTWVTNGAD-----VSGV 470
Qy 541 M-----IFGKE-----SAGASNTALDNVMTITDEBEIKATNPVATE 575
Db 471 RAVRVGYSTDPYGGQOPESDILLRLYSASAEQQNPILN-----AARH 516
Qy 576 RFGTVAVNFQSSSTDPATGDVHAMGALPGMVQDRDVLQGPWAKIPTHGHPHPSPLM 635
Db 517 TFTREARTKLITGSGADGDYKEMWMLPNQWMSAPISRYNPWVKVPRVNRKTLTQD 576
Qy 636 GGFLKNPPQILIK--NTPVPANPPAEFSAFASITQYSTGQSVSELEWELQKNSK 693
Db 577 GSIPMSHPPTGTFIKLARIIPVNGD-----SFLNIYVTGQSVCEVWEVEKRGTK 627
Qy 694 RWNPEVQYTSNYSKANSVD--FTVDNNGLY 721
Db 628 NWRPEYMS-----ATNMSVDATYINNAGVY 653

RESULT 3
COAT_PAVHB STANDARD; PRT; 781 AA.
AC P07299;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE Probable coat protein VP1.
OS Human parvovirus B19.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.
OX NCBI_TaxID=10798;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=Isolate AU;
RX MEDLINE=86200451; PubMed=3701931;
RX Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;
RT "Nucleotide sequence and genome organization of human parvovirus B19
RT isolated from the serum of a child during aplastic crisis."
RL J. Virol. 58:921-936(1986).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```





Db 513 QYNDE-----PNGAIRFTMGYHQGLTTS-----SQLEKRYTFNPQSKCGRAPKQOFNQ 562

Qy 580 VA-VNFOSS-----TDPATG--DVHMGALP-----GMWQORDVYLOQPIWAKIPHT 625

Db 563 QAPLNTLNTGTLPSDPIGGKPMHFMNTLNTYGPLTALNTAPVFPNGQIWDKELDT 622

Qy 626 D--GHFHPSPLMGFGULK-NPPPOILIKNTVPANPPAFESA-TKFASTFYQTQGVSV 681

Db 623 DLKPRHLH---VTAPFVCKNPPGQLFVKIAP---NLTDFFNADSPQOPRIITYSNFWMKG 676

Qy 682 EIEWELOKENSKRWNPVOYTSNTAKSANVDFTVDNNGLYTEPRPIG 728

Db 677 TLTFATKRRSMNMNPIQOHTT-----TAENIGNYL-PTNIG 712

RESULT 5

COAL\_PAVPN STANDARD; PRT; 729 AA.

AC P18546; P22964; Q89816;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Coat protein VP1 [Contains: Coat protein VP2].

OS Porcine parvovirus (strain NADL-2) (PPV).

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.

OX NCBI\_TaxID=10797;

RN [1]

RX MEDLINE=90085785; PubMed=2596019;

RA Vasudevacharya J., Basak S., Srinivas R.V., Companas R.W.;

RT "Nucleotide sequence analysis of the capsid genes and the right-hand

RL terminal palindrome of porcine parvovirus, strain NADL-2.;"

RN [2]

RX MEDLINE=91021005; PubMed=2219713;

RA Vasudevacharya J., Basak S., Srinivas R.V., Companas R.W.;

RT "The complete nucleotide sequence of an infectious clone of porcine

RL parvovirus, strain NADL-2.;"

RN [3]

RX MEDLINE=90010964; PubMed=2794971;

RA Ranz A.I., Mancius J.J., Diaz-Aroca E., Casal J.I.;

RT "Porcine parvovirus: DNA sequence and genome organization.;"

RL J. Gen. Virol. 70:2541-2553(1989).

CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED

CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.

CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT

CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; N32787; AAA46917.1; -

DR EMBL; N32787; AAA46918.1; -

DR EMBL; N38367; AAA46919.1; -

DR EMBL; N38367; AAA46921.1; -

DR EMBL; D00623; BAA00502.1; -

DR PIR; B33302; VCPVPP.

DR PIR; B33743; VCPVNA.

DR HSSP; P30129; 4DPV.

DR InterPro; IPR001403; Parvo coat.

DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein.

FT CHAIN 1 729 COAT PROTEIN VP1.

FT CHAIN 151 729 COAT PROTEIN VP2.

FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DOMAIN 175 190 GLY-RICH.

FT CONFLICT 10 10 G -> GKSPKGVAVILQIIFYITG (IN REF. 3).

FT CONFLICT 16 18 YKY -> TI (IN REF. 3).

FT CONFLICT 56 56 Y -> T (IN REF. 3).

FT CONFLICT 164 164 G -> A (IN REF. 3).

FT CONFLICT 195 195 T -> S (IN REF. 3).

FT CONFLICT 242 242 V -> S (IN REF. 3).

FT CONFLICT 294 294 E -> A (IN REF. 3).

FT CONFLICT 705 705 N -> K (IN REF. 3).

FT CONFLICT 715 715 R -> K (IN REF. 3).

SQ SEQUENCE 729 AA; 80946 MW; EF816E246C80DB42 CRC64;

Query Match 12.1%; Score 482; DB 1; Length 729;

Best Local Similarity 25.5%; Pred. No. 2.8e-22;

Matches 197; Conservative 111; Mismatches 310; Indels 154; Gaps 35;

Qy 43 RGLVLPQYKYLGPNGLDKGEFVNADAAALEHDKAYDQOLKAGDNPYLRYNHADAEFOE 102

Db 9 RGLTLPQYKYLGPNSLDQGEPTNPDAAAKEHDEAYDKYKSGKPNFYFSAADEKPIK 68

Qy 103 RLCEDTSFQNLGRAVFOAKRVLPLGLVEGAKTAPGKRPVQESQBPQSSSGIGK- 161

Db 69 ETEHAKYGGKIGHYFFRAKRAFKL---SETDPTTSQOPEVRRSPRKHGSKPPGKR 125

Qy 162 -----TGQPAKKRLNFGTGSSEVDPDQPLGEPPATPAAVPTTWSAGGGAPMA 212

Db 126 PAPHIFINLAKKKAKGTSNTNSNSNENVEQHPINAGTEL-SATGNESGGGGGG- 181

Qy 213 DNNEGADGVGNAGS--NWHCDSTWLG---RVITTSRTWALPTNNHLYKQIS--SAST 265

Db 182 -GGRGAGGVGVSTGTFFNQTEFOYLGELGVRIITAHASRLIHLNMPHETTKRIHVLNSES 240

Qy 266 GAS-----NDNHFGYSTPMGVDFNFRHCFHSPRDQRLINNNMGFRPKRLNFKLFNIQV 321

Db 241 GVAGQMVQDDAHTQMTVPMSLIDANAGVWFNPADQQLISNNMTEINLVSPEQEIFNVVL 300

Qy 322 KEVT---TNDGVTTIANLSTVOVSDSEYQLPVLGSAHQGCLPPFPADVFMIPQYGY 378

Db 301 KTITESATSPPTKIYNNDLTASLMAVLDTNNTLPTTPAAPRSETLGFYPLPTKPTQYRY 360

Qy 379 L-----TLNNGSOAVGRS-----SFYCLE-YFPSSOMLRTGNPF-TFSYTFE 417

Db 361 YLSCIRNLNPPYTGQSQITDSIQTGLHSDIMPTTIENAVPIHLLRTGDEFSTGIYHFD 420

Qy 418 EVPFHSSVAHSQSILDR-----LNMPLI--DOYLYLARTQNGSQAQKOLLFSRGS 467

Db 421 TKPL--KLTHSQWNRSLGLPPKLLTEPTTEGDQHPGLPAANTRKGVHQTINNSYTEAT 478

Qy 468 PAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNTWTGASKYNLNGRESIINP--GTAMA 525

Db 479 ----AIRP-----AOQYNTPTMNPFSYNGGGPF-----LTPIVPTADT 512

Qy 526 SHKDDKFFPMSGVMIFGKESAGASNTALD---NVMTDDEEIK--ATNP-----VAT 574

Db 513 QYNDDPEV-----GAIRFTMDYQHGHLTTSQLELRYTFNPQSKCGRAPK 557

Qy 575 ERFGTVA--VNFQSSS-----TDPATG--DVHMGALP-----GMWMDORDVYLOQPIWA 620

Db 558 QQFNQOAPLNLNTNNGTLLPSDDPIGGKSNMHFMNTLNTYGPLTALNNTAPVFPNGQIWD 617

Qy 621 KIPHTD--GHFHPSPLMGGFGLK-NPPPOILIKNTPVPANPPAFESA-TKFASTFYQTQST 676

Db 618 KELDTDLKPRHLH---VTAPFVCKNPPGQLFVKIAP---NLTDFFNADSPQOPRIITYSN 671

QY 677 GQVSVBIEWELQKENSXRMNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIG 728  
 Db 672 FWWKGLTFTAKRSMNMNPIQHTT-----TAENIGNYI-PTNIG 712

## RESULT 6

ID COAL\_PAVPK STANDARD; PRT; 729 AA.

AC P52501;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coat protein VP1 [Contains: Coat protein VP2].  
 OS Porcine parvovirus (strain Kresse) (ppv).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=73487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96183900; PubMed=8642680;  
 RA Bergeron J., Hebert B., Tijssen P.;  
 RT "Genome organization of the Kresse strain of porcine parvovirus;  
 RT Identification of the allotropic determinant and comparison with  
 RT those of NADL-2 and field isolates.";  
 RL J. Virol. 70:2508-2515(1996).

CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
 CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.  
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U44978; AAC40230.1; -;  
 DR EMBL; U44978; AAC40231.1; -;  
 DR HSSP; P30129; 4DPV

DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein.

FT CHAIN 1 729 COAT PROTEIN VP1.  
 FT CARBOHYD 151 729 COAT PROTEIN VP2.  
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 729 AA; 80835 MW; B6345BFA0568A1F6 CRC64;

Query Match 12.0%; Score 480; DB 1; Length 729;

Best Local Similarity 25.6%; Pred. No. 3.7e-22;

Matches 196; Conservative 112; Mismatches 315; Indels 144; Gaps 34;

QY 43 RGLVPGYKYLPGNLDKGEVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEPQE 102

Db 9 RGLTPLYKYLPGNSLDQGEPTNPSDAAAKEHDEAYDKYKGNPKYFYFSADEKFIK 68

QY 103 RLQEDTSFGNIGRAVFOAKRVLPLGLVEGAKTAPGKRVEQSPQEPDSSSIGK- 161

Db 69 ETEHAKDYGGKIGHYFRAKRAFAPKL---SETDPTTSQQPEVRSPRKHGSKPPGKR 125

QY 162 -----TGQOPAKRLNFGQTGDSSEVPDPQPLGEPDPATPAAVGPTTMASGGGAPMA 212

Db 126 PAPRHIFINLAKKAKGTSTNSNSMSENVEQHNPNAGTEL-SATGNSGGGGGGG--- 181

QY 213 DNNEGADGVGNASGNWH--CDSTWLGD---RVITTSRTWALPTYNHLYKQIS--SAST 265  
 Db 182 -GGRGAGGVSTGSGFNQTEFYQLGEGVLRTATASRLIHLNMPHEHYTKIHLVNSGS 240  
 QY 266 GAS-----NDNHFGYGTGWPYDFNRFRCHSPRQWQRLINNNWGRPKRLNFKLPNIQV 321  
 Db 241 GVAGQMVQDDAHTQMTVTEWSLIDANAWGVWFPADWQLISNNMTBILNVSFQEIFNVVL 300  
 QY 322 KEVT---TNDGVTTIANNLTSTVQVFSDEYOLPYVLGSAHOGCLPPRPADVPMPYOYCY 378  
 Db 301 KTIITESATSPPTKIYNNDLTASLWALDVTNNLTPTTPAARSETLGFYPMWPKTKPTQYRY 360  
 QY 379 L-----TLNNGSQAAGRS-----SFYCLE-YFPSSQMLRTGNF-TFSYTFE 417  
 Db 361 YLSCTRNLPPTTGTGSOQITDSIQTGLHSDIMFYTIENAVPIHLRLTGDEFSTGIYHFD 420  
 QY 418 EYVPHSSVAHSQSLDR-----LMNPLI--DOVLYLNRNTQNSGSAQNKDLFSPRGS 467  
 Db 421 TKPL--KLTHSWQTNRSLSGLPPKLLTEPTTEGDOHPGTLPAANTRKGVHQTINNSYTEAT 478  
 QY 468 PAGMSVQPKWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMA 525  
 Db 479 ----AIRP-----AQGVNTPYNNFYSNGGPF-----LTPIVPTADT 512  
 QY 526 SHKDEDEKFFPMGVMIFKESAGASANTALDNVMTDDEEIKATNP-----VATERFGT 579  
 Db 513 QYNDDE---PNGAIRFTMGVYHGQLTTS-----SQELERYTFNFPQSKGRAPKQFNO 562  
 QY 580 VA-VNFSQSS-----TDPATG--DVHANGALP-----GMWQDRDVLQGIWAKIPIHT 625  
 Db 563 QAPLNLNTNNTLLPSDPDGKPMHFMNTLNTYGPULTALNNTAPVFPNGQIWKDELDT 622  
 QY 626 D--GHFHPSPLMGGGLK-NPPQILIKNTVPANPPAEFSA-TKFASPTOYSTQGVSV 681  
 Db 623 DLKPLRH--VTAPFVKNNPPGQLFVKIAP---NLTDNFADSPQOPRIIYSNFWPWKG 676  
 QY 682 EIEWELQKENSXRMNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIG 728  
 Db 677 TLTPTAKRSMNMNPIQHTT-----TAENIGNYI-PTNIG 712

## RESULT 7

ID COAT\_FPV

AC P04864; Q65112; STANDARD; PRT; 727 AA.

DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Coat protein VP1 [Contains: Coat protein VP2].

OS Feline panleukopenia virus (FPV).

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.

OX NCBI\_TaxID=10786;

RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=85265017; PubMed=2991581;

RA Carlson J., Rushlow K., Maxwell I., Maxwell F., Winston S., Hahn W.;

RT "Cloning and sequence of DNA encoding structural proteins of the

RL J. Virol. 55:574-587(1985).

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M10824; AAA47161.1; -;

DR EMBL; M10824; AAA47162.1; -;

DR FIR; A03701; VCPVIF.

DR HSSP; P30129; 4DPV.

DR InterPro; IPR001403; Parvo\_coat.  
KW Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 727  
FT CHAIN 144 727  
FT CARBOHYD 168 168  
FT CARBOHYD 190 190  
FT CARBOHYD 207 207  
FT CARBOHYD 323 323  
FT CARBOHYD 586 586  
FT CARBOHYD 648 648  
FT CARBOHYD 660 660  
FT CARBOHYD 165 180  
SQ SEQUENCE 727 AA; 80344 MW; 77E0F6F554C0C6E CRC64;  
  
Query Match 11.1%; Score 444.5; DB 1; Length 727;  
Best Local Similarity 23.6%; Pred. No. 5.5e-20;  
Matches 187; Conservative 117; Mismatches 295; Indels 189; Gaps 35;  
  
QY 43 RGLVLPGYKYLPGFNLGDKGEPVNAADAALEHDKAYDQQLKAGDNPYLRYNHADAEPQE 102  
DB 10 RGLVPPGYKYLPGFNLGDKGEPVNAADAALEHDKAYDQQLKAGDNPYLRYNHADAEPQE 69  
  
QY 103 RLOEDTSFGNIGRAVFOAKRVLPLGLVEBGAKTAPGKRPVQSPQEPSSSGIGKTG 162  
DB 70 QTKDATDWGKGIGHYFFRAKKAIAPIVLTDPDHPSTSRPTK-PTKRSKPPHIFINLAK- 127  
  
QY 163 GQPAKRLNFGQTGDSSEVP---DPOPLGEPFA--TPAAVGPPTMASGGGAPMADNNE 216  
DB 128 -----KKKAGAGQVRDNLAPMSDGAQVDPGQPAVRNERATGSGNGSGGG-----GG 176  
  
QY 217 GADGVNAGSNWH-----CDSTWLGDRVITTTSTRTWALPTYNHLYKQI-----SS 262  
DB 177 GSGVGISTGTENNQTEPFLENGWV--EITANSRLVHLNMPESYKRVVNNMDKTA 234  
  
QY 263 ASTGASNDNHFGYSTPGYFDNFRPHCHFSPRDQRLNNWGFPRKLNFKLNIQVK 322  
DB 235 VGNMALDDTHVQIVTPWLSVDANAWGVNFWNPGDQWLIVNTWSELHLSVFEQEIFNVLK 294  
  
QY 323 EVT---TNDGVTTIANLNTSTVQVSDSEYQLPVYLGSAHQGL-----PPPPA----- 368  
DB 295 TVSBESATQPTKVYNNDLTASLWALDLSNTHPTPAAMRSETLGFYMPKPIPTPMRY 354  
  
QY 369 ---DVPIMPOYGYLTLNNGSQAVGRS-----SFYCLE-YFPQMLRTGNPF-TFS 413  
DB 355 FQWDRTLIPSH-----TGTSGTPTNIYHGTDPDDVQFYTIENSVPVHLLRTGDEFATGT 408  
  
QY 414 YTFEEVPFHSSVAHSQSLDLN--PLIDQYLYLNRTOQSQAQNKDLFSRGSAGM 471  
DB 409 FFDCKP--CRLTHTWQTRALGLPP-----FLNSLPCSEGATNFGDIGVQDKRRGV 459  
  
QY 472 S-----VQPKN-WLPGPCYRQVRSK-----TKTDNNNSNFTWT 504  
DB 460 TQMGNTDIYATEIMPAEYGSAPYSFEASTQGFPIAAGRGAQTENQAA---D 516  
  
QY 505 GASKY---NUNGRESINPGT-----AMASHKDEDEKFFPMGSGVMIPGKESAGASNTALD- 556  
DB 517 GDRYAFGRHQGQKTTTGTETPERFTYIAHQDT-----GRYPAGDWIQNIF 563  
  
QY 557 NVMITDDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHANGALPGMWQDRDVLQ 616  
DB 564 NLPVTNDNVLPTDPIG---GKTGINY--TNIPNTYGLTALNVV-----PVYNG 610  
  
QY 617 PIWAKIPTHDTGHFPHSLMGGFLK-----NPPQILIKNTFVPAN-----PPA 660  
DB 611 QIWKEFDTD-----LKPRLHVNAFPVQNNCPGQLFKVAPNLNTNEYDPA 657  
  
QY 661 EFSATKFAEITQYSTGQSVSELEWLOKENSKRNPVEQYTSNYAKSANVDFTVNNG 720  
DB 658 SANMSR-----IVTYSDFWKGKLVFAKLRASHTWNPDIQMSIN-----VDNQFN 703  
  
QY 721 YTEPRPIG 728

DB 704 YL-PNNIG 710  
  
RESULT 8  
ID COAT\_MEVA STANDARD; PRT; 722 AA.  
AC P27437;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Mink enteritis virus (strain Abashiri) (MEV).  
OC Viruses; ssRNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10793;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91202123; PubMed=2016597;  
RA Kariatsunari T., Horiuchi M., Hama E., Yaguchi K., Ishiguro N.,  
RA Goto H., Shinagawa M.;  
RT "Construction and nucleotide sequence analysis of an infectious DNA  
RT clone of the autonomous parvovirus, mink enteritis virus.";  
RL J. Gen. Virol. 72:867-875(1991).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
CC EMBL; D00765; BAA00666.1; -;  
DR PIR; B38350; VCPVME.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 722  
FT CHAIN 139 722  
FT DOMAIN 160 177  
FT CARBOHYD 163 163  
FT CARBOHYD 185 185  
FT CARBOHYD 202 202  
FT CARBOHYD 318 318  
FT CARBOHYD 581 581  
FT CARBOHYD 643 643  
FT CARBOHYD 655 655  
SQ SEQUENCE 722 AA; 79823 MW; 9DADBCAB2EF9622 CRC64;  
  
Query Match 11.0%; Score 437.5; DB 1; Length 722;  
Best Local Similarity 23.6%; Pred. No. 1.5e-19;  
Matches 186; Conservative 117; Mismatches 295; Indels 189; Gaps 35;  
  
QY 44 GLVLPGYKYLPGFNLGDKGEPVNAADAALEHDKAYDQQLKAGDNPYLRYNHADAEPQE 103  
DB 6 GLVPPGYKYLPGFNLGDKGEPVNAADAALEHDKAYDQQLKAGDNPYLRYNHADAEPQE 65  
  
QY 104 LQEDTSFGNIGRAVFOAKRVLPLGLVEBGAKTAPGKRPVQSPQEPSSSGIGKTG 163  
DB 66 TKDATDWGKGIGHYFFRAKKAIAPIVLTDPDHPSTSRPTK-PTKRSKPPHIFINLAK-- 122  
  
QY 164 QPAKRLNFGQTGDSSEVP---DPOPLGEPFA--TPAAVGPPTMASGGGAPMADNNE 217  
DB 123 -----KKKAGAGQVRDNLAPMSDGAQVDPGQPAVRNERATGSGNGSGGG-----GG 172  
  
QY 218 ADGVNAGSNWH-----CDSTWLGDRVITTTSTRTWALPTYNHLYKQI-----SSA 263  
DB 173 GSGVGISTGTENNQTEPFLENGWV--EITANSRLVHLNMPESYKRVVNNMDKTA 230  
  
QY 264 STGASNDNHFGYSTPGYFDNFRPHCHFSPRDQRLNNWGFPRKLNFKLNIQVK 323  
DB 231 KGNMALDDTHVQIVTPWLSVDANAWGVNFWNPGDQWLIVNTWSELHLSVFEQEIFNVLK 290



```

611 QIWDKEPDTD-----LKPRLVHVNAPFVQCNNCPGQLFVKVAPNLTWEYDPA 657
661 EFSATKPFASFTQYSTQGVSEIWELOKENSKEWNEVQYTSNYAKSANDVFTVNNGL 720
658 SANMSR-----IVTYSDFWFKGLVFKAKLRASHWTNPIQOMSIN-----VDNQFN 703
721 YTEPRPIG 728
704 YV-PNNIG 710

RESULT 10
ID COAT_PAVCB STANDARD; PRT; 727 AA.
AC Q11213.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Canine parvovirus (strain B) (CPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=59284;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91272479; PubMed=1647068;
RA Parrish C.R.;
RT "Mapping specific functions in the capsid structure of canine parvovirus and feline panleukopenia virus using infectious plasmid clones.";
RT Virology 183:195-205(1991).
RL CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, OF A COMBINATION OF VP2, VP3, AND SOME VP1.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
DR EMBL; M38245; AA02799.1; -
DR EMBL; M38245; AA02800.1; -
DR HSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 727
FT CHAIN 144 727
FT CARBOHYD 168 168 (POTENTIAL).
FT CARBOHYD 190 190 (POTENTIAL).
FT CARBOHYD 207 207 (POTENTIAL).
FT CARBOHYD 323 323 (POTENTIAL).
FT CARBOHYD 586 586 (POTENTIAL).
FT CARBOHYD 648 648 (POTENTIAL).
FT CARBOHYD 660 660 (POTENTIAL).
SQ SEQUENCE 727 AA; 80342 MW; 115F3E1A79098EBE CRC64;

Query Match 10.8%; Score 430.5; DB 1; Length 727;
Best Local Similarity 23.2%; Pred. No. 4e-19;
Matches 184; Conservative 112; Mismatches 297; Indels 201; Gaps 32;

QY 43 RGLVLPGYKLPNGLDGEPVNAADAAALSHDKAYDQQLKAGNPVLYRNHDAEFQE 102
DB 10 RGLVFPYKILGPNLSLDQGETNFSDAAKHDEAYAYLRSGKNPYLYFSPAQRFID 69
QY 103 RLQEDTSFGCNLGRAVFOAKRVLPLGLVEBGAKTAFGKRPVQSPQEPDSSSIGIKT 162
DB 70 QTKDAKDWGKGIGHYFFRAKAIAPVLTDPDHPSTSRPTK-PTKRSKPPPHIFINLAK- 127
QY 163 GQOPAKKRLNFGQTGDSSEVP-----DPQLGEPFA--TPAAVGPMTMASGGAPMADNNE 216

```

```

128 -----KKKAGAGQVKRDNLAPMSDGAQVQPDGGQPAVNERATSGNGSGGG-----GG 176
217 GADGVGNASQNW-----CDSTWLDGVRVITTTSTWALPTNNHLYK01-----SS 262
177 GSGVGLSTGTFNQTQEFKLENGWV--ETANSSRLVHLNMPESENYRVRVNNMDKTA 234
263 ASTGASNDNHYFGYSTPWGYDFENRFCHFSPRDQWRLINNNMGFRPKRLNFKLFNIOVK 322
235 VNGNMALDDIHAQIVTPWSLVANANGVMFNPQDWQILVNTMSELHLVSEQEIFNVVLK 294
323 EVT---TNDGVTTIANNLTSTQVFSDEYQLPYVLGSAHQGL-----PPPA----- 368
295 TVSESATQPPTKVYNNDLTASLMVALDSNNTMPTPAAMRSETLGFVPKWTPTTPWRY 354
369 ---DVFMIPOYGYLTLNNGSOAVGRS-----SFYCLE-YFPSQMLRTGNF-TFS 413
355 PQWDRTLIPSH-----TGSGTPTIYHCTDPPDVQFYTIENSVPVHLLRTGDEFATGT 408
414 YTFEEVFPFHSSYAHSSQSLDLAMN--PLIDQYLYLANRTQNSGSAQNKDQLLFSRGSAGM 471
409 FFDCKE--CRLTHTMTQNEALGLPP-----FLNSLPQSEGATNFGDI-----GV 451
472 SVQPKWLPQPCYRQQRVSKTKDNNNSNF-----TWGASKYNLNGRESIINP-- 520
452 -----QDDKREGVTQMGNTNYITEATIMRPAEYGVSAFYSEASTQGPFK 497
521 ---GTAMASHKDDKFFPMGVMIFGKESAGASNTALD----- 556
498 TPIAAGRGGAQTENQAADGNPRYAFGRHQGKTTTGTETPERFTYIAHQDTGRYPGDM 557
557 -----NVMITDEEIKATNPVATERFGTVAVNPFQSSSTDPATGDVHAMGALPGMWQDR 610
558 IQNINFLNPLVNDNVLLPTDPIG---GKTGINY--TNIENTYGLPLTALANVP----- 604
611 DVYLOGPIWAKIPHTDGHFHPSPLMGGFGLK-----NPPQILLIKNTVPAN 657
605 PVPNGQIWDKEPDTD-----LKPRLVHVNAPFVQCNNCPGQLFVKVAPNLTN 651
658 ---PPAFSFSATKPFASFTQYSTQGVSEIWELOKENSKEWNEVQYTSNYAKSANDVFT 714
652 EYDPAASANMSR---IVTYSDFWFKGLVFKAKLRASHWTNPIQOMSIN----- 697
715 VDNNGLYTEPRPIG 728
698 VDNQFNIV-PSNIG 710

RESULT 11
ID COAT_MUMIM STANDARD; PRT; 718 AA.
AC P07302; Q9WMH2; Q9WMH3;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Murine minute virus (strain MMV1) (Murine parvovirus).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10795;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86115415; PubMed=3502703;
RA Astell C.R., Gardiner E.M., Tattersall P.;
RT "DNA sequence of the lymphotropic variant of minute virus of mice, MVM(i), and comparison with the DNA sequence of the fibrotropic prototype strain.";
RT J. Virol. 57:656-669(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85242059; PubMed=3855242;
RA Sahli R., McMaster G.K., Hirt B.;
RT "DNA sequence comparison between two tissue-specific variants of the autonomous parvovirus, minute virus of mice.";
RT

```

Nucleic Acids Res. 13:3617-3633(1985).

[3]

RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718.

RA Llamas-Saiz A.L., Agbandje-Mckenna M., Wlkoiff W.R., Bratton J.,

RA Tattersall P., Roessmann M.G.;

RT "Structure determination of Minute Virus of mice.";

RL Acta Crystallogr. D 53:93-100(1997).

CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X02481; CAB46507.1; -.

DR EMBL; X02481; CAB46508.1; -.

DR EMBL; M12032; AAA69569.1; ALT\_INIT.

DR PIR; B23008; VCPVIM.

DR PDB; 1MVM; 25-PEB-98.

DR InterPro; IPR001403; Parvo\_coat.

DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein; 3D-structure.

FT CHAIN 1 718 COAT PROTEIN VP1.

FT CHAIN 132 718 COAT PROTEIN VP2.

FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 502 502 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 635 635 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 671 671 N-LINKED (GLCNAC. .) (POTENTIAL).

FT DOMAIN 157 172 GLY-RICH.

FT STRAND 144 144 A -> G (IN REF. 2).

FT STRAND 182 186

FT STRAND 191 204

FT STRAND 213 217

FT TURN 220 221

FT TURN 228 231

FT STRAND 235 245

FT TURN 251 253

FT HELIX 256 265

FT STRAND 266 269

FT STRAND 274 289

FT STRAND 296 301

FT TURN 303 304

FT STRAND 307 307

FT STRAND 309 312

FT HELIX 322 325

FT TURN 326 326

FT TURN 333 334

FT STRAND 337 338

FT STRAND 340 345

FT STRAND 351 351

FT TURN 357 358

FT STRAND 367 367

FT TURN 373 375

FT HELIX 381 384

FT STRAND 390 390

FT STRAND 396 397

FT STRAND 408 409

FT STRAND 428 428

FT STRAND 438 438

FT HELIX 442 444

FT TURN 458 460

FT STRAND 466 467

FT TURN 473 474

FT STRAND 477 477

FT STRAND 484 484

FT TURN 494 496

FT TURN 499 503

FT STRAND 509 509

TURN 512 513

STRAND 525 525

HELIX 534 536

STRAND 546 546

TURN 557 558

STRAND 559 559

TURN 575 577

TURN 583 584

STRAND 585 586

STRAND 600 600

STRAND 618 618

STRAND 627 630

STRAND 636 636

STRAND 649 649

STRAND 652 668

STRAND 683 683

STRAND 691 691

STRAND 692 694

SEQUENCE 718 AA; 79017 MW; B43C1762ED5F74B9 CRC64;

Query Match 10.6%; Score 424.5; DB 1; Length 718;

Best Local Similarity 23.7%; Pred. No. 9.2e-19;

Matches 184; Conservative 104; Mismatches 298; Indels 189; Gaps 31;

QY 45 LVLPQYKYLGPFGNGLDKGPEPVNAADAAALBHDKAYDQOLKAGDNPYLRYNHADADEFQRL 104

DB 1 MYPGYKYLGPFGNGLDKGPEPVNAADAAALBHDKAYDQOLKAGDNPYLRYNHADADEFQRL 60

QY 105 QEDTSFGNGLGRAVFOAKRVLEPLGLVEGAKTAPGKKRVEQSPQSPDSSSGIGKTG- 163

DB 61 KDAKDWGKGVGHYFRTKRAFAPKLTADSE-----PG-----TSGVSRAGK 101

QY 164 -----QQPAKKRLNFG---QTGDSSEVPDPOLG-----EPPATPAVAGPTTM 203

DB 102 RTRPAYIFINQARAKKLTSSAAQQSSQTGDSQPDGNAVHSAARVERAADGPG-- 159

QY 204 ASGGGAPMADNNEGADGVGNAGNWHCHDS--TWLGD---RVITTTSTRTWALPTYNHLYK 258

DB 160 GSGGG-----GSGGGGVGVSTGYDNQTHYRFLGDGWEITALATRLVHLMNPSENYC 213

QY 259 QI-----SSASTGASNDNHFGYSTPWGFDNRRFHCHFSRDWQRLINNNGWFRPKR 311

DB 214 RIRVHNTTDSVKGNNMAKDDAHEQIWTWPSLVDANAMGWLQPSDMQYICNTMSQLNLS 273

QY 312 LNFKLENIQKVEVTND---GVTTIANLNTSTVQVPSDSEVQLPYVLGSAHQGCLPFP 367

DB 274 LDQEIFNVLTVTVEQDGGQAIKIYNDLTACHMVAVDNNILPYTPAASMETLGFYP 333

QY 368 ADVFMIPOGY-----LTLNN-----GSQAVGRSSFCYCLEYFPPS-QMLRTG 407

DB 334 WKPTIASPYRYFCVDRDLSVTYENQEGTIEHNVNGTPEKGNSSQFFTIENQITLLRTG 393

QY 408 NNF-TFSYTPPEVPHSSYAHQSGLDRLMN--PLIDQLYLYLNRTQNSGSAQNKDILFS 464

DB 394 DEFATGYTYPDTNPFV--KLTHWTQNRQLGQPPLLSTF-----PEADTDAGT-----LTA 441

QY 465 RGSFAGMSVQPKNWL-----PGPCYRQQRVSKTKTDNNNSN 500

DB 442 QGSRHGATQMEVNVVSEAIRTPAQVGFQCPHNDFEASRACP-FAAPKVPADVTCGVDR 500

QY 501 FTWTGASKYNLNGRESIINPGCTAMASHKDDDKF-----PMSGVMIFGKESAGAS 551

DB 501 ANGSVRSYSGKHGHNMAAHGPAERYVTWDTNFGSGRDRTRDGFQISAPLVVPPPLNGIL 560

QY 552 NTALDNVMTDBEERIKATNPVATPFGTVAVNPOS--SSTDPAICGDVAMGALPGMWQD 609

DB 561 TNA-----NPIGTKN-----DIHPSNVFNSYGLPTAFSH-----P 590

QY 610 RDVYLQGPWAK---IPHTDGHFPHSPLMGGFGLKNPPPIILIKNTP--VPANPPAFPSA 664

DB 591 SPVYFQGIWDKELDLEHKPRLHITAPFV---CKKNAPGQMLVRLGPNLTQYDENGATL 647

QY 665 TKFASFITQYSTGTQVSVBEIWELOKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 719







QY 327 ----NDGVTTIANNLTSTVQFSDSEYQLPVYLSAQHQCGL-----PPFPAD---VFMI 373  
Db 291 QGAGODALKVYNDLITACMVALDSNNILPYTPAAQTSETLGFYFQWKPATAPRYFYFEM 350  
QY 374 POYGYLTUNGSQ-----AVG-----RSSFYCLE-YFPQMLRTGNFTF-SYTFBEVP 420  
Db 351 PQLSVTSNSAEGTQITDTTIGEPOALNSQFFTIENLTPLTLRLTGDEFTTGTTFIENDP 410  
QY 421 PHSSVA-----HSQSLDRLMN-PLIDQVLYLNRTONQSGSAQNKDLIF----- 463  
Db 411 LKLTHTWQNRHLACLOGITDLPTSDTATASLTANGDRFGSTQTONVNVYVTEALTRPAQ 470  
QY 464 -----SRGSPAGMSVQPKWLPFGPCYRQORVSKTKTDNNNSNFTWTGASKYNL 511  
Db 471 IGFQPHDNFEANRGPFKVPVP-----LDITAGEDHDAN-----GAIRENY 513  
QY 512 NGR--ESIINPCTA-----MASHKDDDEKFFPMGSMVMI FGKESAGASNTALDNVM 559  
Db 514 GKQHGDMWAKGAAPERYTWDAIDSAGRDRTARCFV-----QSAPISIPPNQOI 563  
QY 560 ITDDEEIKATNPVATEREGTVAVNFQSSSTDPATCDVHAMGALPGMWQDRDVYLOQPI 618  
Db 564 LQREDALAGRTWHTYTNVNSYGLSAPHPDP-----IYPNGOI 603  
QY 619 WAK---IPHTDGHFHPSPLMGFGKLNPPPIILIKNTVPVNPAPF--SATKPFASFITQ 673  
Db 604 WDKELDLEKPLRLHTAPFV---CKNNPPQGLFVHLGP---NLTDQDFNSTTVSRIVT- 656  
QY 674 YSTGVSVIEIWELOKENSIRNPEVQYTSN 704  
Db 657 YSTFYWKGLPKAKLRPNLTWNPVYQATD 687

RESULT 15

COAT\_PAVC7  
ID COAT\_PAVC7 STANDARD; PRT; 722 AA.  
AC P04863;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Canine parvovirus (strain 780929) (CPV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10789;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85185696; PubMed=3989914;  
RA Rhode S.L. III;  
RT "Nucleotide sequence of the coat protein gene of canine parvovirus.";  
RL J. Virol. 54:630-633(1985).  
RN [2]  
RP REVISIONS.  
RA Rhode S.L. III;  
RL Submitted (JUL-1989) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 228-473 FROM N.A.  
RX MEDLINE=86098657; PubMed=3942033;  
RA Parrish C.R., Carmichael L.E.;  
RT "Characterization and recombination mapping of an antigenic and host range mutation of canine parvovirus.";  
RL Virology 148:121-132(1986).  
CC -I- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC OF A COMBINATION OF VP2, VP3, AND SOME VP1.  
CC -I- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----

DR EMBL; M10989; AAA47149.1; -.  
DR EMBL; M11871; AAA47147.1; -.  
DR PIR; A03702; VCPVCN.  
DR HSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 722 COAT PROTEIN VP1.  
FT 139 722 COAT PROTEIN VP2.  
FT CARBOHYD 163 722 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 185 185 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 202 202 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 581 581 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 643 643 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 655 655 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT DOMAIN 160 181 GLY-RICH.  
FT CONFLICT 409 409 K -> R (IN REF. 3).  
FT CONFLICT 438 439 AT -> DI (IN REF. 3).  
FT CONFLICT 447 447 P -> Q (IN REF. 3).  
FT CONFLICT 451 451 K -> R (IN REF. 3).  
FT CONFLICT 454 454 V -> I (IN REF. 3).  
SQ SEQUENCE 722 AA; 79826 MW; BBB37380B4371429 CRC64;  
  
Query Match 9.7%; Score 386.5; DB 1; Length 722;  
Best Local Similarity 23.3%; Pred. No. 2e-16;  
Matches 176; Conservative 111; Mismatches 300; Indels 201; Gaps 32;  
  
QY 49 GYKLGPFNGLDKGEVPVNAADAALEHDKAYDQOLKAGDNPYLRYNHADAEFORLQEDT 108  
Db 11 GYKLGPKSLDQGEPTNPSDAAKEHDEAYAAVLRSGKPYLYFSPADQRFIDQTKDAK 70  
QY 109 SFGNGLGRAVFOAKRVLEPLGLVEGAKTAPGKKRPVEGSPQBPDSGSGIGTKGQQPAK 168  
Db 71 DWGCKIGHYFPRACKATAPVLTDPDHPSTSRPTK-PTKRSKPPPHIFINLAK-----K 123  
QY 169 KRLNFGQTGDSSEVP-----DQPLGEPPA--TPAAGVPTTMASSGGAPMADNNGADGVG 222  
Db 124 KKAGAGQVKRDNLAPMSDAGVQDGGQPAVNERATSGSGGGGG-----GGSGGVG 177  
QY 223 NASGNWH-----CDSTWLGDRVITSTRTWALPTNNHLYKQI-----SSASTGAS 268  
Db 178 ISTGTFFNQTEFKLENGW--EITANSSRLVHLNMPSEKDRRVVNNMDKTAVNGMA 235  
QY 269 NDHNYFGYSTPMGYFDFNRHFCHSPRDMQRLNINNWGFRPKRLNFKLFINQVEVT--- 325  
Db 236 LDDIHAQIVTPWLSLVANAMGWENFGDQWLIVNTMSLHVSFEQEIFNVVLKTVSSA 295  
QY 326 TNDGVTTIANLSTVQVFSSEYQLPVYLSAQHQCGL-----PPFPA-----DVP 371  
Db 296 TOPPTKVYNNDLTASLMVALDSNNMTPTTPAMRSETLGFYFQWKPATAPRYFYFEM 355  
QY 372 MIPQGYLTUNGSQAVGRS-----SFYCLE-YFPQMLRTGNFTF-SYTFBEVP 419  
Db 356 LIPSH-----TGTSGTPTNIYHGTDPDDVQFYTIENSVPVHLRLTGDEFTTGTFFDCK 409  
QY 420 PFHSSYAHQSQSLDRMN--PLIDQVLYLNRTONQSGSAQNKDLIFSGSGPAGMSVQPKN 477  
Db 410 P--CRLTHTWQNRHLGPP-----FLNSLPOSEGATNFGDI-----GVP----- 447  
QY 478 WLPGPCYRQORVSKTKTDNNNSNF-----TWGASKYNLNGRESIINP-----GTA 523  
Db 448 -----QKKGRTVQMGNTNITEATIMRPAEYGSAPYYSFEASTQGFPTPIAAG 498  
QY 524 MASHKDDDEKFFPMGSMVMI FGKESAGASNTALD----- 556  
Db 499 RGGAGTDENQAADGNPRVAFGRQHGKKTITTTGETPERFTYIAHQDTGRYPEGDMTQINFP 558  
QY 557 NMWITDEEIKATNPVATEREGTVAVNFQSSSTDPATCDVHAMGALPGMWQDRDVYLOQ 616  
Db 559 NLPVTNDNVLLPIDPIG---GKTGINY--TNIFNTYGLPTALNNVP-----PVYNG 605  
QY 617 PIWAKIPTHDTGHFHPSPLMGFGKLNPPPIILIKNTVPVNPAPF--SATKPFASFITQ 660

```
Db      606 QIWDKEFTD-----LKPRLHVNAPFVCONNCPGOLFVKLAPNLNTNEYDPDA 652
Qy      661 EPSATKFAFITYSTGQVSVEIEWELOKENSKRWNPEVQYTSNYAKSANVDFTVDNNGL 720
Db      653 SANMSR-----IVTYSDFWWKGLVFKAURASHWTWNPICOMSGIN-----VDNQFN 698
Qy      721 YTEPRPIG 728
Db      699 YV-PSNIG 705
```

Search completed: January 21, 2004, 16:02:10  
Job time : 14.6014 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:01:27 ; Search time 16.9331 Seconds  
(without alignments)  
1839.046 Million cell updates/sec

Title: US-09-807-802a-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWNLSGIR.....NNGLYTEPRICRYLTRPL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3402.5	85.3	735	US-09-321-589-1	Sequence 1, Appli
2	2486.5	62.3	734	US-09-532-594B-4	Sequence 4, Appli
3	1930.5	45.9	598	US-09-532-594B-16	Sequence 16, Appli
4	1690.5	42.4	544	US-09-532-594B-18	Sequence 18, Appli
5	1410	35.3	756	US-09-438-268-4	Sequence 4, Appli
6	479.5	12.0	543	US-08-856-841-22	Sequence 22, Appli
7	430.5	10.8	500	US-08-856-841-16	Sequence 16, Appli
8	430.5	10.8	501	US-08-856-841-18	Sequence 18, Appli
9	429.5	10.8	486	US-08-856-841-19	Sequence 19, Appli
10	326	8.2	415	US-08-856-841-20	Sequence 20, Appli
11	318	8.0	395	US-08-856-841-13	Sequence 13, Appli
12	310	7.8	264	US-08-856-841-14	Sequence 14, Appli
13	303	7.6	398	US-08-856-841-21	Sequence 21, Appli
14	284	7.1	387	US-08-856-841-17	Sequence 17, Appli
15	216.5	5.4	579	5223424-13	Patent No. 5223424
16	192	4.8	584	US-09-022-949-2	Sequence 2, Appli
17	142	3.6	210	US-08-856-841-9	Sequence 9, Appli
18	142	3.6	227	US-08-856-841-15	Sequence 15, Appli
19	142	3.6	250	US-08-856-841-12	Sequence 12, Appli
20	124	3.1	3060	US-08-487-826B-14	Sequence 14, Appli
21	121	3.0	2736	US-09-252-991A-30227	Sequence 30227, A
22	119.5	3.0	1394	US-08-296-791-2	Sequence 2, Appli
23	119.5	3.0	1394	PCT-US95-10661A-2	Sequence 2, Appli
24	119	3.0	1095	US-09-107-532A-3855	Sequence 3855, Ap
25	114.5	2.9	434	US-08-710-249-4	Sequence 4, Appli
26	114.5	2.9	434	US-09-220-157A-4	Sequence 4, Appli
27	113.5	2.8	655	US-08-469-202-27	Sequence 27, Appli

ALIGNMENTS

RESULT 1  
US-09-321-589-1  
; Sequence 1, Application US/09321589  
; Patent No. 6498244  
; GENERAL INFORMATION:  
; APPLICANT: PATEL, SALIL D.  
; APPLICANT: MCARTHUR, JAMES G.  
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS  
; FILE REFERENCE: 39672  
; CURRENT APPLICATION NUMBER: US/09/321.589  
; CURRENT FILING DATE: 1999-05-28  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: Adeno-associated virus  
US-09-321-589-1

Query Match 85.3%; Score 3402.5; DB 4; Length 735;  
Best Local Similarity 83.3%; Pred. No. 1.5e-289;  
Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY	1	MAADGYLPDWLNLSGIR	REWWDLKPGAPKPKANQKQDDGRGLVLP	GYKYLGPNGLD	60	
DB	1	MAADGYLPDWLNLSGIR	REWWDLKPGAPKPKANQKQDDGRGLVLP	GYKYLGPNGLD	60	
QY	61	KGEPVNAADAAALEHDKAYDQOLKAGDN	PYLRYNHADAEFQERLQEDTSFGN	LGRAVFQ	120	
DB	61	KGEPVNAADAAALEHDKAYDQOLKAGDN	PYLRYNHADAEFQERLQEDTSFGN	LGRAVFQ	120	
QY	121	AKKRVLEPLGLVEGAKTAPGKRPVEQS	POEPDSSSGIGTKGQOPAKKRLNFG	QTGDSSE	180	
DB	121	AKKRVLEPLGLVEEVPKTA	PGKRPVEHS	PVEPDSSSGTGKAGQOPAKKRLNFG	QTGDAD	180
QY	181	SVDPDQPLGEPAPAAVGPPTTMA	SGGAPMADNNEGADGVNAGSNHCH	DSMTLGDRI	240	
DB	181	SVDPDQPLGQPPAAPSGLGT	TMTATGSGAPMADNNEGADGVNAGSNHCH	DSMTMGDRI	240	
QY	241	TTSTRTWALPTYNHLYKQISSAST	GASNDNHYFCYSTPWGYFDFNRFCH	FSPRDWORL	300	
DB	241	TTSTRTWALPTYNHLYKQISSQS	-GASNDNHYFCYSTPWGYFDFNRFCH	FSPRDWORL	299	
QY	301	INNNGGPRKRLNFKLFNI	QVKEVTNDGVTTIANNLTSTVQV	FTDSEYQLPYVLSAHQ	360	
DB	300	INNNGGPRKRLNFKLFNI	QVKEVTQNDGTTIANNLTSTVQV	FTDSEYQLPYVLSAHQ	359	
QY	361	GCLPPFPADVPMI	POGYLTLNNGSOAGRSSFCYCLEY	FPSOMLRTGNNTFSY	TFEEVVP	420
DB	360	GCLPPFPADVPMI	POGYLTLNNGSOAGRSSFCYCLEY	FPSOMLRTGNNTFSY	TFEEDVP	419

```
QY 421 FHSSYAHQSLSLRLNPLIDQYLYLNRTQNSGSAQNKDLLFSRGSPAGMSVQPKWLP 480
Db 420 FHSSYAHQSLSLRLNPLIDQYLYLNRTQNSGSAQNKDLLFSRGSPAGMSVQPKWLP 479
QY 481 GPCYQQRVSKTKTNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMGVS 540
Db 480 GPCYQQRVSKTKSADNNNSEYSGTATKYLHNGRDSLVPNGPAMASHKDDKDFPMGVS 539
QY 541 MIFGKESAGASANTALDNVMTDBEEIKATNPVATERFGTVAVNFOSSTDPATGDVHAMG 600
Db 540 LIFGKQSEKTNVDEIKWMTDBEEIRTNVATERFGTVAVNFOSSTDPATGDVHAMG 599
QY 601 ALPGMWQDRDYLQSPWAKIPHTDGHFHPSPLMGGFGLKXPPQILIKNTPPVPPANPPA 660
Db 600 VLPGMWQDRDYLQSPWAKIPHTDGHFHPSPLMGGFGLKXPPQILIKNTPPVPPANPPA 659
QY 661 EFSATKFASTFYQSTGQSVSEIWELOKENSCKRWNPVEQYTSNYAKSANVDPTVNNGL 720
Db 660 TFSAAKFASTFYQSTGQSVSEIWELOKENSCKRWNPVEQYTSNYAKSANVDPTVNNGL 719
QY 721 YTEPRPIGTRYLTRPL 736
Db 720 YSEPRPIGTRYLTRNL 735

RESULT 2
US-09-532-594B-4
; Sequence 4, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP1
US-09-532-594B-4

Query Match 62.3%; Score 2486.5; DB 4; Length 734;
Best Local Similarity 63.5%; Pred. No. 3-2e-209;
Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;

QY 4 DGYLPDWNLENTSEGIREWMDLKPAGPKPKANQKQDDGRGLVLPYKYLGPNGLDKGE 63
Db 3 DGYLPDWNLENTSEGIREWMDLKPAGPKPKANQKQDDGRGLVLPYKYLGPNGLDKGE 62
QY 64 PVNAADAALAEHDKAYDQQLKAGDNVLYRNHADAESFQRLQEDTSFGNGLGRAVFOAKK 123
Db 63 PVNAADAALAEHDKAYDQQLKAGDNVLYRNHADAESFQRLQEDTSFGNGLGRAVFOAKK 122
QY 124 RVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGTKGQOPAKKRLNF-GOTGDSSEV 182
Db 123 RVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGTKGQOPAKKRLNF-GOTGDSSEV 182
QY 183 PDQPILEPPATPAAVPTTMAAGGAPMADNNEGADGVGNAGSNHWCDSWTWLGDRVITTT 242
Db 183 PEGSTSG-----AMSDSEMAAAGAAVEGGQADGVGNAGSNHWCDSWTWLGDRVITTT 236
QY 243 STYTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYDFNRFCHFSPRDWQRLIN 302
```

```
Db 237 STRTWALPTYNNHLYKRLGE-----SLOSTNYNGESTPMGYDFNRFCHFSPRDWQRLIN 292
QY 303 NNWFRPKRLNFKLNFQVKEVTINDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQC 362
Db 293 NNWGRPKAMRVKLFNFQVKEVTSNGETTVANNLTSTVQIFADSSYELPYVMDAGQSG 352
QY 363 LPPEPADVFMIPQYGY---LTLNNGSAQVGRSSFCYCLEYFPSPQMLRTGNNTFTSYTPEV 419
Db 353 LPPEPNDVFMVPOYGYCGLVTGNTSQQTDRNAPCYCLEYFPSPQMLRTGNNTFTSYTPEV 412
QY 420 PFHSSYAHQSLSLRLNPLIDQYLYLNRTQNSGSAQNKDLLFSRGSPAGMSVQPKWLP 475
Db 413 PFHSSYAHQSLSLRLNPLIDQYLYLNRTQNSGSAQNKDLLFSRGSPAGMSVQPKWLP 469
QY 476 KMWLPGPCYQQRVSKTKTNNNSNFTWTGAS---KY-----NLNGRESIINPGTAMASHK 528
Db 470 KMWLPGPSIKQGGFSKTA--NQYKIPATGSDLSLKIVETHSTLDGRWSALTGPPMATAG 527
QY 529 DDEKFFPMGSMVIFGKESAGASANTALDNVMTDBEEIKATNPVATERFGTVAVNFQSS 588
Db 528 PADSK-FSNSQLIFAGPKQNGTATVPCTLIFTSSEELATNATDMDMGNLPGDQDQNS 586
QY 589 TDPATGDVHAMGALPGMWQDRDYLQSPWAKIPHTDGHFHPSPLMGGFGLKXPPQIL 648
Db 587 NLPTVDRLTALGAVPMWQNRDIYQGPWAKIPHTDGHFHPSPLMGGFGLKXPPQIL 646
QY 649 IKNTPPVPPANPPAEFSATKFASTFYQSTGQSVSEIWELOKENSCKRWNPVEQYTSNYAKS 708
Db 647 IKNTPPVPPANPPAEFSATKFASTFYQSTGQSVSEIWELOKENSCKRWNPVEQYTSNYAKS 706
QY 709 ANVDFTVNNGLYTEPRPIGTRYLTRPL 736
Db 707 NSLLWAPDAAGKYTEPRAIGTRYLTRPL 734

RESULT 3
US-09-532-594B-16
; Sequence 16, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP2
US-09-532-594B-16

Query Match 45.9%; Score 1830.5; DB 4; Length 598;
Best Local Similarity 57.8%; Pred. No. 7.9e-152;
Matches 355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;

QY 138 TAPGKKRPVEQSPQEPDSSSGIGTKGQOPAKKRLNF-GOTGDSSEVPPQPLGEPATPA 196
Db 1 TAPGKKRPVEQSPQEPDSSSGIGTKGQOPAKKRLNF-GOTGDSSEVPPQPLGEPATPA 196
QY 197 AVGPPTTMAAGGAPMADNNEGADGVGNAGSNHWCDSWTWLGDRVITTTSTRTWALPTYNNH 256
Db 55 MSDDSEMAAAGAAVEGGQADGVGNAGSNHWCDSWTWLGDRVITTTSTRTWALPTYNNH 114
```



121 AKRVRLEPLGLVEGAKTAPGKRPVEQSPQEPSSGIGKTGQOPAKKRLNFQGTGDS 180  
121 AKRVRLEPLGLVEGAKTAPGKRPVEQSPQEPSSGIGKTGQOPAKKRLNFQGTGAD 180  
181 SVDPQPIGEPATPAAVGPPTMASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVI 240  
181 SVDPQPIGEPATPAAVGPPTMASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVI 240  
241 TSTRWALPTNNHLYKQISSASTGASND-----NHFGYSTPWCYFDPNRFHC 290  
240 CTFGRQFLIPYDPEHHYGFSPFASSCHNAGKAVCTISPIMGYSTPWRILDFNALNL 299  
291 HFSRDMORLINNNMGFRPKLNFLEIYQKEVT--TNDGVTTIANLNTSTVQVFSDE 348  
300 FFSPLFQHLIENYGSIAPDALTVTISEIAVKDVTDKTGGV-QVTDSTTGRCLWLDHE 358  
349 YQLPYVLSAHQGLPPPPADVFIPQYGLTLNN-GSQAVG-----RSSFYCLEY 398  
359 YKYPVYLGQGGDTLAPLPIWYPPQVAYLTGVDVNTQGISGDSKKLASEESAFYVLEH 418  
399 FPSQMLRTGNNFTSYTFEYVPHSSYAHQSOLDRLMPLIDQYLYLNRNTQNSGSAQN 458  
419 SSFQLGTGGTATMSYKFPVPPENLEGCQHFVEMYNPL---YGRSLGVDPDTLGGDPKF 475  
459 KDLFSRGSAGMSYQPKNMLPGPCYRQORVSKTKTDNNNSFTWTGASKYNLAGRESII 518  
476 RSL-----THEDHAIQPNFPGPLVNSVSTKEGDSNTGAKALTGLSTGTSQNTRI 530  
519 NPG-TAMASHKDDKDFPMGSGWMLFGKESAGASNTALDNV-----MITDBEIKATN 570  
531 RGPVSPQYHHWDTKYVTGYNALSHGQTTYG---NAEDKEYQGVGRFPNEKEQLKQL 587  
571 PVATERFCTVAVNFSSSTDPATGVDVHAGALPGMWQDRDVLQGPITWAKIPHTDGHF 630  
588 GINMHTY-----FNNKGTQYTDIE-RPLMVGSVNRRALHYESQLWSKIPNLDSPK 640  
631 PS-PLMGFGKLNPPQILKNTVPANPPAFSAFATKPFITQYSTQGVSEIWEWEL-Q 688  
641 TOPAALGKWLHQPPIFLK--ILPQSGPIGKSGMIGITLVAQVAVGIMTWTFTKLP 698  
689 KENSKRWNP 698  
699 RKATGRWNPQ 708

## RESULT 6

US-08-856-841-22  
; Sequence 22, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: BROOKS HAIDT HAPFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658

; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 543  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PEPTIDE  
; DESCRIPTION: N/A  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
; POSITION IN GENOME: GENETICALLY ENGINEERED PEPTIDE  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.  
; AUTHORS: FIELD, A.M.  
; AUTHORS: CANT, B.  
; AUTHORS: WIDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET  
; VOLUME: I  
; ISSUE:  
; PAGES: 72 - 73  
; DATE: 1975  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 22:  
US-08-856-841-22

Query Match 12.0%; Score 479.5; DB 3; Length 543;  
Best Local Similarity 27.1%; Pred. No. 2.1e-33;  
Matches 141; Conservative 79; Mismatches 236; Indels 75; Gaps 16;  
QY 212 ADNNEGADGVGNAS--GNWHCDSTWLGDRVITSTRWALPTNNHLYKQISSASTGASN 269  
DB 16 ABASTGAGGGSNSVKSWMSEGATFSANSVTCFSRQFLIPYDPEHHYKVPSPAASCHN 75  
QY 270 D-----NHVFGYSTPWCYFDPNRFCHSPRDWQRLNNWGFPRKRLNFKLPNI 319  
DB 76 ASGKEAKVCTISPTMGYSTPWRILDFNALNLFSPLEFQHLIENYGSIAPDALTVTISEI 135  
QY 320 QYKEVT--TNDGVTTIANLNTSTVQVFSDEYQLPYVLSAHQGLPPFPADVFMIPOYG 377  
DB 136 AKVDVTDKGGV-QVTDSTTGRCLWLDHEIKYTPYVLGQGGDTLAPLPIWYTPPQYA 194  
QY 378 YLTINN-GSQAVG-----RSSFYCLEYFSPQMLRTGNNFTSYTFEVPFHSYAH 427  
DB 195 YLTVDVNTQGISGDSKKLASEESAFYVLEHSSFQLGTGTSMSYKFPVPPENLEGC 254  
QY 428 QSGLRLMNLIDQYLYLNRNTQNSGSAQNKOLLFSRGSAGMSVQPKWLPFGPCTRQQ 487

Db 255 SOHFYEMNPL---YGRSLGVPDTLGGDPKFRSL-----THEHAIQPNFMGPPLVNSV 306  
Qy 488 RVSKTKDNNNSFTWTGASKYNLNGRESIING-TAMASHKDEDEKFPFMSGVMIFGKE 546  
Db 307 STKEGSSNTGAKALTGLSTGTSQNTSRISLRPGVPSPYHHWDTDKYVTGINAISHGQT 366  
Qy 547 SAGASNTALDNV-----MITDEEIKATNPVATERFGTVAVNFOSSTDPATGDVHAM 599  
Db 367 TYG---NAEDKEYOQGVGRFPNEKEQLKQLGLNMHTY-----FPMKGTQOYTDQIE-R 416  
Qy 600 GALPGMWQORDVYLOGPIWAKIPIHTDGHFHS-PLMGFGELKNPPQILIKNTPVPANP 658  
Db 417 PLMVGSVNNRRALHYESQLMSKIPNLDSDSKTQFAALGGWGLHQPQQI----- 465  
Qy 659 PASFSATKPSFITQSTGVSVIEIWEIWL-QKENSKEWNP 698  
Db 466 -----FLKQVAGIMVTMTFKLGPRTKATGRWNQ 495

RESULT 7  
US-08-856-841-16  
: Sequence 16, Application US/08856841  
: Patent No. 6274307  
: GENERAL INFORMATION:  
: APPLICANT: ERWIN SOUTSCHEK  
: APPLICANT: MANFRED MOTZ  
: TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
: TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
: NUMBER OF SEQUENCES: 28  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
: STREET: 99 PARK AVENUE  
: CITY: NEW YORK  
: STATE: NY  
: COUNTRY: USA  
: ZIP: 10016  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 3.5" FLOPPY DISC  
: COMPUTER: AT&T - IBM COMPATIBLE  
: OPERATING SYSTEM: MS-DOS Version 6.2  
: SOFTWARE: ASCII  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/856,841  
: FILING DATE:  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/08/214,658  
: FILING DATE: 16-MARCH-1994  
: APPLICATION NUMBER: US 07/917,096  
: FILING DATE: 4-AUGUST-1992  
: APPLICATION NUMBER: PCT/DE91/00106  
: FILING DATE: 8-FEBRUARY-1991  
: APPLICATION NUMBER: DE40038262  
: FILING DATE: 8-FEBRUARY-1990  
: ATTORNEY/AGENT INFORMATION:  
: NAME: ROBINSON, WILLIAM R.  
: REGISTRATION NUMBER: 27,224  
: REFERENCE/DOCKET NUMBER: LKR-9222-A  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (212) 697-3355  
: TELEFAX: (212) 557-5635  
: TELEX: NONE  
: INFORMATION FOR SEQ ID NO: 16:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 500  
: TYPE: AMINO ACID  
: TOPOLOGY: LINEAR  
: MOLECULE TYPE:  
: DESCRIPTION: PEPTIDE  
: HYPOTHETICAL: N/A  
: ANTI-SENSE: N/A  
: FRAGMENT TYPE: INTERNAL

: ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
: IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
: POSITION IN GENOME: N/A  
: FEATURE:  
: NAME/KEY: N/A  
: LOCATION: N/A  
: IDENTIFICATION METHOD: amino acid analysis and  
: IDENTIFICATION METHOD: mass spectrometry  
: OTHER INFORMATION:  
: PUBLICATION INFORMATION:  
: AUTHORS: COSSART, Y.E.  
: AUTHORS: FIELD, A.M.  
: AUTHORS: CANT, B.  
: AUTHORS: WIDDOWS, D.  
: TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
: JOURNAL: LANCET  
: VOLUME: 1  
: ISSUE:  
: PAGES: 72 - 73  
: DATE: 1975  
: DOCUMENT NUMBER:  
: FILING DATE:  
: PUBLICATION DATE:  
: RELEVANT RESIDUES IN SEQ ID NO: 16:  
: PUBLICATION INFORMATION:  
: AUTHORS: MANIATIS, T.  
: AUTHORS: FRITSCH, E.F.  
: AUTHORS: SAMBROOK, J.  
: TITLE: MOLECULAR CLONING  
: JOURNAL: COLD SPRING HARBOR, NY  
: VOLUME:  
: ISSUE:  
: PAGES: 1982  
: DATE: 1982  
: DOCUMENT NUMBER:  
: FILING DATE:  
: PUBLICATION DATE:  
: RELEVANT RESIDUES IN SEQ ID NO: 16:  
: PUBLICATION INFORMATION:  
: AUTHORS: SMITH, D.B.  
: AUTHORS: JOHNSON, K.S.  
: TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
: TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
: TITLE: GLUTATHIONE S. TRANSFERASE  
: JOURNAL: GENE  
: VOLUME:  
: ISSUE: 67  
: PAGES: 31 - 40  
: DATE: 1988  
: DOCUMENT NUMBER:  
: FILING DATE:  
: PUBLICATION DATE:  
: RELEVANT RESIDUES IN SEQ ID NO: 16:  
: US-08-856-841-16  
: Query Match 10.8%; Score 430.5; DB 3; Length 500;  
: Best Local Similarity 29.5%; Pred. No. 3.7e-29;  
: Matches 122; Conservative 45; Mismatches 170; Indels 77; Gaps 10;  
Qy 47 LPGYKYLGPFGNLDKGEFVNADAAALHDKAYDQOLKAGDNPYLRYNKADAEFQERLOE 106  
Db 133 LFGTYVGPNGNELQAGPPQSAVDSAAIRHDFRYSQAKLGINPYTHWTVADEELLKNIGN 192  
Qy 107 DTSGFNGNIGRAVFOAKKRVLEPLGLVBEGAKTAPCKKRPVEQSPOEPDSSSGIGKTGOOP 166  
Db 193 ETGFOAQVVKDYF-----TLKGAAPVAHFQ-----GSLPEVP 225  
Qy 167 AKKRLNFGQTGDSSEVPDPQPLGEPPATPAVGPPTTMASGGGAPMADNNEGADGVGNASG 226  
Db 226 AYNA-----SEKYPSTSVNSAEASTGA-----GGG-----GNSVKS 258  
Qy 227 NWHCDSTWLGDRVITTTSTRTWALPTYNNHLYKQISSASTGASND-----NHYFGY 276



Db 259 MWSEGATFSANSVTCFTRQRLPYDPEHHYKVFSPAASSCHNASKGKAEKVCCTISPINGY 318  
Qy 277 STPWGYDPNRFPHCFSPRDWQRLNNWGRPKRLNFKLFNIOVKET--TNDGVTTIA 334  
Db 319 STPWRYLDFNALNLFSPLEFQHLIENYGSAPDALTTVISEIAVKDVTDTGGGV-QVT 377  
Qy 335 NNLSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPQGYLTLAN-GSQAVG----- 389  
Db 378 DSTTGRCLMLVDHEYKYPYVLGQGGDTLAPELPIWVYPPQYAYLTVGDVNTQGISGDSK 437  
Qy 390 -----RSSFCYCLEYPPSQMLRTGNFTSYTPEEYPPHSSVAHSQSGLRLMPL 438  
Db 438 KLAESAFYVLEHSSFOLLGTGTSMSYKFPVPVPPENLEGCSQHFEYMTNPL 491

RESULT 8

US-08-856-841-18  
; Sequence 18, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016

COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 501  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE:  
; DESCRIPTION: PEPTIDE  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
; POSITION IN GENOME: N/A

FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.  
; AUTHORS: FIELD, A.M.  
; AUTHORS: CANT, B.  
; AUTHORS: WIDDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET  
; VOLUME: I  
; ISSUE:  
; PAGES: 72 - 73  
; DATE: 1975  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 18:  
US-08-856-841-18

Query Match 10.8%; Score 430.5; DB 3; Length 501;  
Best Local Similarity 29.5%; Pred. No. 3.7e-29;  
Matches 122; Conservative 45; Mismatches 170; Indels 77; Gaps 10;  
Qy 47 LPGYKYLGPFGNGLDKGBFVNAADAAALHDKAYDQQLKAGDNPYLRYNHADAFOERLQE 106  
Db 134 LPGTNYVGPNGNELQAGPPQSAARIHDFRYSQLAKGINPYTHWTVADDELLKNIKN 193  
Qy 107 DTSGGNLGRAVFOAKKRVLEPLGLVEGAKTAPKKRPVEQSPQEPDSSSGIGKTGOOP 166  
Db 194 ETGFAQVVKDYF-----TLKGAAPVAHFQ-----GSLPEVP 226  
Qy 167 AKRLNFGQTGDSSEVPDPQPLGEPPATPAAGVTPTMASGGCAPMADNNEGADGVGNASG 226  
Db 227 AYNA-----SEKPSMTSVNSAEASTGA-----GGG-----GSNSVKS 259  
Qy 227 NWHCDSTWLGDRVITTTTRTWALPTYNNHLYKQISSASTGASND-----NHFGY 276  
Db 260 MWSEGATFSANSVTCFTRQRLPYDPEHHYKVFSPAASSCHNASKGKAEKVCCTISPINGY 319  
Qy 277 STPWGYDPNRFPHCFSPRDWQRLNNWGRPKRLNFKLFNIOVKET--TNDGVTTIA 334  
Db 320 STPWRYLDFNALNLFSPLEFQHLIENYGSAPDALTTVISEIAVKDVTDTGGGV-QVT 378  
Qy 335 NNLSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPQGYLTLAN-GSQAVG----- 389  
Db 378 DSTTGRCLMLVDHEYKYPYVLGQGGDTLAPELPIWVYPPQYAYLTVGDVNTQGISGDSK 438  
Qy 390 -----RSSFCYCLEYPPSQMLRTGNFTSYTPEEYPPHSSVAHSQSGLRLMPL 438  
Db 438 KLAESAFYVLEHSSFOLLGTGTSMSYKFPVPVPPENLEGCSQHFEYMTNPL 492

RESULT 9

US-08-856-841-19  
; Sequence 19, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 577-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: I  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 19:  
US-08-856-841-19  
Query Match 10.8%; Score 429.5; DB 3; Length 486;  
Best Local Similarity 28.8%; Pred. No. 4.3e-29;  
Matches 120; Conservative 45; Mismatches 169; Indels 83; Gaps 7;  
QY 47 LPGYKYLPGFNLGKGEFVNAADAALEHDKAYDQOLKAGDNPYLYRYNHADAEFQERLQ 106  
DB 125 LPGTNYVPGNLTQAGPQSAVDAAIHDFRYSQAKLGINPYTHWTVADELLKNKN 184  
QY 107 DTSPGNGNLGRAVFOAK--KRVLEPLGLVEEAKTAPGKKRPVEQSPQSDSSGIGKTG 163

185 ETGFAQVVDYFTLKGAAADVAHFQGLSEVPAYNASEKYPMSVNSAEASTGAGGG 244  
164 QOPAKRLNFCQTGDSSEVPDPQLGEPPTAAVGPPTMASGGAPMADNNEGADGVGN 223  
245 SNPKSM----- 251  
224 ASGNWCHDSTWLGDRVITTTSTRTWALPTNNHLYKQISSASTGASND-----NH 273  
252 ---WSEGATFSANSVCTFSRQFLIPYDPEHYKVFSPAASSCHNAGSKAKVCTISPI 307  
274 FGSTPMGYFDNRFPHCFSPRDMQRLNNWGRPKRLNFKLNIQKVT--TNDGVT 331  
308 MGYSTPMRYLDFNALNLFSPLEFQHLTIENYGSTAPDALTVTISEIAVKDVTDKTGGV- 366  
332 TIANNLTSTVOFSDSEYQLPVLGSAHQGLCPPADVEMIPOGYLTLANN-GSOAVG- 389  
367 QVTSATGRCLMVDHEYKPYVLGGQDQTLAPLPIWVYPPQYAYLTVGDVNTQGISG 426  
390 -----RSSFYCLEYPPSQMLRTGNNFTSYTFEEVPPHSSVAHSQSGLRLMPL 438  
427 DSKKLASESAFYVLEHSSFOLLGTGGTATWSYKFPVPPENLEGCSQHFYEMYNPL 483

RESULT 10  
US-08-856-841-20  
Sequence 20, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 415  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE





DATE: 1982  
DOCUMENT NUMBER: US/08/214, 658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
PUBLICATION RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
US-08-856-841-14

Query Match 7.8%; Score 310; DB 3; Length 264;  
Best Local Similarity 31.3%; Pred. No. 4.9e-19;  
Matches 78; Conservative 33; Mismatches 104; Indels 34; Gaps 6;

QY 194 TPAAVGPTTASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYN 253  
DB 5 TFSLAEAESTGAGG-----GNSVKSMSWSEGAFTSANSVTCFTRQFLIPYDP 53  
QY 254 NHLYKOISASTGASND-----NHVFGYSTPWGYFDNRFCHFSPRDMQRLINN 303  
DB 54 EHHYKVFSPAASSCHNASGKAQVCTISPTMGYSTPWRYLDFNALNLFSPLEFQHLIEN 113  
QY 304 NMGFRPKRLNFKLFNIOVKEVT--TNDGVTTIANLTSVQVFSQYOLPYVLGSAHQG 361  
DB 114 YGSIAPDALTVTISEIAVDVDTKGGV-QVTDSTTGRCLMLVDHXYKPYVLGQGQDT 172  
QY 362 CLPPFPADVFMIQYGYLTINN-GSQAVG-----RSSFYCLEYFSPQMLRTGNFT 411  
DB 173 LAPFLPIWYVFPQYAYLTVGDVNTQGISGDSKCLASESAFVLEHSSFQLLGTGGTAS 232  
QY 412 FSYTPPEVP 420  
DB 233 MSYKFPVP 241

RESULT 13  
US-08-856-841-21  
Sequence 21, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRES:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856, 841  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214, 658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCI/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 398  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PEPTIDE  
DESCRIPTION: N/A  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.B.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE: 72 - 73  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 21:  
US-08-856-841-21

Query Match 7.6%; Score 303; DB 3; Length 398;  
Best Local Similarity 25.4%; Pred. No. 3.9e-18;  
Matches 96; Conservative 48; Mismatches 168; Indels 66; Gaps 12;

QY 346 DSEYQLPVVLGSAHQGCLPPFPADVFMIQYGYLTINN-GSQAVG-----RSSFYC 395  
DB 4 DHEYKPYVLGQGDQLAPFLPIWYVFPQYAYLTVGDVNTQGISGDSKCLASESAFV 63  
QY 396 LEYFSPQMLRTGNFTFSYTFEEVFPFHSSYAHQSGLDRMLDQYLYLNRTONQSGS 455  
DB 64 LEHSSFQLLGTGTATMSYKFPVPPENLEGSCSHFYEMYNPL---YGSRLGVPDTLGGD 120  
QY 456 AQNKDLLFSRGSAGMSVQPKNWLPGCVQRQVRVSKTKTDNNNSNFTWTGASKYNLNGRE 515  
DB 121 PKFRSL-----THEDHAIQPNFMFGLVNSVSTKEGSSNTGACKALTGLSTGTSTQNR 175  
QY 516 STINFG-TAMASHKDEDEKFFPFMSGVMI FGKESAGASNTALDNV-----MITDEEEK 567  
DB 176 ISLRFGPVSPYHHWDTKYVTGINAISHGQTTYG---NAEDKEYQGVGRFPNKEQLK 232

QY 568 ATPVATERFGTVAVNFQSSSTDPATGDVHAMGALFGVMQDRDVLQGPWAKIPHTDG 627  
Db 233 QLQGLNHTY-----FPNKGTOQYTDQIB-RPLMVGSVNRRALHYESQLNSKIPNLD 285  
QY 628 HFHPS-PLMGFGKXPPQILIKNTPVPANPPAERFASATKFSFIQYGTGQVSVEIWE 686  
Db 286 SPTQPAALGGWGLHQPPOI-----FKYHKVGQLEVLNQWE 323  
QY 687 LQENSK-----RWN 697  
Db 324 LLPPNPWELQLHNGP 341

## RESULT 14

US-08-856-841-17  
; Sequence 17, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 387  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE:  
; DESCRIPTION: PEPTIDE  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
; POSITION IN GENOME: N/A  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A

; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.  
; AUTHORS: FIELD, A.M.  
; AUTHORS: CANT, B.  
; AUTHORS: WIDDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET  
; VOLUME: 1  
; ISSUE:  
; PAGES: 72 - 73  
; DATE: 1975  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 17:  
; PUBLICATION INFORMATION:  
; AUTHORS: MANIATIS, T.  
; AUTHORS: FRITSCH, E.F.  
; AUTHORS: SAMBROOK, J.  
; TITLE: MOLECULAR CLONING  
; JOURNAL: COLD SPRING HARBOR, NY  
; VOLUME:  
; ISSUE:  
; PAGES:  
; DATE: 1982  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 15:  
; PUBLICATION INFORMATION:  
; AUTHORS: SMITH, D.B.  
; AUTHORS: JOHNSON, K.S.  
; TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
; TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
; TITLE: GLUTATHIONE S. TRANSFERASE  
; JOURNAL: GENE  
; VOLUME:  
; ISSUE: 67  
; PAGES: 31 - 40  
; DATE: 1988  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 17:  
; US-08-856-841-17

Query Match 7.1%; Score 284; DB 3; Length 387;  
Best Local Similarity 28.0%; Pred. No. 1.7e-16;  
Matches 82; Conservative 28; Mismatches 119; Indels 64; Gaps 6;  
QY 47 LPGYKYLGFNGLDKGEPVNAADAALEHDKAYDQDLKAGDNPYLRYNHADAERLOE 106  
Db 133 LFGTYVYVGNELQAGPPQSAVDSAAIRHDFRYSQAKLGINPYHTWTVADELLKNKN 192  
QY 107 DTSGGNLGRAVFOAKKRVLEPLGLVBEGAKTAPCKRPVQSPOEPDSSSGIGKTGOOP 166  
Db 193 ETGFOAQVVKDYF-----TLKGAAPVAHFQ-----GSLPEVP 225  
QY 167 AKKRLNFGQTDGSESVDPQPLGEPPTAAVGPPTMASGGGAPMADNNEGADGVGNASG 226  
Db 226 AYNA-----SEKYPSMTSVNSAEASTGA-----GGG-----GNSVKS 258  
QY 227 NWHCDSTWLGDRVITTTTRTWALPTYNHLYKQISSASTGASND-----NHYFGY 276  
Db 259 MMSEGATFSANSVTCTFSRQFLIPYDPEHHYKVFSPAASCHSNASGKEAKVCTISPIMGY 318  
QY 277 STPMGYFDNRFCHFSRDRQLNNNNWGRPKELNFKLFNIQVKEVTNDG 329  
Db 319 STPMRYLDFNALNLFPSPLEFQHLIENTYGSIAPODALTVTISEIAVKDVTDTKG 371

Qy 695 WNEVQYTSYAKSANYDFTVDNNGLYTERPPIG 728  
Db 540 WNPQQHTT-----TAENIRKYI-PTNIG 562

Search completed: January 21, 2004, 16:08:00  
Job time : 18.9331 secs

RESULT 15  
5223424-13  
;PATENT NO. 5223424  
;APPLICANT: COCHRAN, MARK;CHIANG, CHRISTINA H.;MACDONALD,  
;RICHARD D.  
;TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND  
;HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO  
;ACID SEQUENCE  
;NUMBER OF SEQUENCES: 16  
;CURRENT APPLICATION DATA:  
;APPLICATION NUMBER: US/07/225,032  
;FILING DATE: 27-JUL-1988  
;PRIOR APPLICATION DATA:  
;APPLICATION NUMBER: 78,519  
;FILING DATE: 27-JUL-1987  
;APPLICATION NUMBER: 933,107  
;FILING DATE: 20-NOV-1986  
;APPLICATION NUMBER: 902,887  
;FILING DATE: 02-SEP-1986  
;APPLICATION NUMBER: 887,140  
;FILING DATE: 17-JUL-1986  
;APPLICATION NUMBER: 823,102  
;FILING DATE: 27-JAN-1986  
;APPLICATION NUMBER: 773,430  
;FILING DATE: 06-SEP-1985  
;SEQ ID NO:13:  
;LENGTH: 579  
5223424-13

Query Match 5.4%; Score 216.5; DB 6; Length 579;  
Best Local Similarity 22.6%; Pred. No. 2.8e-10;  
Matches 143; Conservative 84; Mismatches 250; Indels 157; Gaps 33;

Qy 179 SESVDPQPLGPPATPAAGVPTTMAAGGAGMADNNEGAGVGNAGNWH--CDSTWLG 236  
Db 2 SENVEQHINAGTEL-SATGNESGGGGG-----GGRGAGGVGVSTGSFNNQTEFYILG 56

Qy 237 D---RVITSTRTWALPTNNHLYKQIS--SASTGAS-----NDNHVFGYSTPMGYDFNR 287  
Db 57 EGLVRITAHASRLIHLNMPHEHYKRIHVLNSESAGVAGQVQDDAHTQMTVPWSLIDRNA 116

Qy 288 FHCHSPROMORLINNWGFRPKLNFKLFNIQVKEVT---TNDGVTTIANNLTSTVOVF 344  
Db 117 WGVNFPADWDQISNNMTINLVSFEQEIFNVVLKTTITESATSPPSKIYNNDLTASLMVA 176

Qy 345 SDEYOLPYVLGSAHQCLPPPPADVFMIPQYGYL-----TLNNGSOAVGRS--- 391  
Db 177 LDTNNTLPTTPAARSETLGFYPLWPKTYRYYLSCIRNLNPPYTGQSQQITDSIQT 236

Qy 392 -----SFYCLE-YFYSQMLRTGNF-TFSYTFEEVPFHSSYAHQSQSLDRMLNPLIDQYL 443  
Db 237 GLHSDIMFYTIENAVPIHLRTGDEFSTGIYHFDTKPL--KLTHSWQTNR----- 284

Qy 444 YYLNRTONQSGSAQNKDLLFSRSGPAGMSVQP-----KWLPGPCYRQ---QRVSKTK 493  
Db 285 -----SGLPPKVLTEPTTEGDQHPGTLPGANTRKGYHQITNNSY 324

Qy 494 TD-----NNNSFTWTGASKYNLNGRESIINP--GTAMASHKDDDEKFFPMG 539  
Db 325 TEATLRPAQGVNTPYMNFYSNGGPF-----LTPIVPTADTQYYDDE---PNGA 372

Qy 540 VNIPIKESAGASNTALDNVMTDEEIKATNP-----VATERPGTVA-VNFQSSS----- 588  
Db 373 IRFTMGYQHGHLTTS-----SQELERYTFNQSKGRAPKQKQFNQQAFLNLTNNGTL 426

Qy 589 --TDPATG--DVHAMGALP-----GMVQDRDVLQGPWAKIPIHTD--GHEHPSPLMG 637  
Db 427 LPSDPIGGKSNKHFMTLNTYGLTALNNTAPVFPNGQIWDKELDTDLKPLRH---VTAP 483

Qy 638 FGLK-NPPFQILIKNTFPVNPANPAEFSA--TKFASFTQYSTQGSVEIEWELQKENSKR 694  
Db 484 FVCKNNPPGOLFVKIAP---NLTTDDFNADSPQOPRIITD-SNFWKKGTLTFTAKMRSSNM 539

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 15:59:46 ; Search time 36.5361 Seconds  
(without alignments)  
2602.280 Million cell updates/sec

Title: US-09-807-802a-15  
Perfect score: 3251  
Sequence: 1 TAPGKRPEVQSPQDPDSS.....NNGLYTEPRPICTRYLTRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3251	100.0	599	21 AAY711168	Adeno-associated v
2	3251	100.0	736	21 AAY711167	Adeno-associated v
3	3229	99.3	736	22 AAB59847	AAV6 capsid protei
4	2906	89.4	534	21 AAY711169	Adeno-associated v
5	2832	87.1	736	22 AAB59846	AAV3B capsid prote
6	2815	86.6	736	22 AAB59845	AAV3A capsid prote
7	2764.5	85.0	735	22 AAG65792	Adeno-associated v
8	2764.5	85.0	735	22 AAM51508	Adeno-associated v
9	2764.5	85.0	735	22 AAB59844	AAV2 capsid protei

10	2764.5	85.0	735	23 AAU98974	Adeno-associated v
11	2759.5	84.9	598	22 AAG65793	Adeno-associated v
12	2759.5	84.9	598	22 AAM51509	Adeno-associated v
13	2759.5	84.9	598	23 AAU98975	Adeno-associated v
14	2738	84.2	734	22 AAB50326	Adeno-associated v
15	2481.5	76.3	533	22 AAG65794	Adeno-associated v
16	2481.5	76.3	533	23 AAU98976	Adeno-associated v
17	2477.5	76.2	533	22 AAM51510	Adeno-associated v
18	1830.5	56.3	598	19 AAM46313	AAV4 VP2 coat prot
19	1830.5	56.3	598	24 AAG73940	Adeno-associated v
20	1830.5	56.3	734	19 AAM46308	AAV4 VP1 capsid pr
21	1830.5	56.3	734	24 AAG73935	Adeno-associated v
22	1719	52.9	736	18 AAM19000	Duck parvovirus ca
23	1701	52.3	732	16 AAR85385	Barbary duck parvo
24	1700.5	52.3	588	21 AAY58161	Adeno associated v
25	1700.5	52.3	588	23 AAU11406	Adeno-associated v
26	1700.5	52.3	724	21 AAY58160	Adeno-associated v
27	1700.5	52.3	724	23 AAU11405	Adeno-associated v
28	1690.5	52.0	544	19 AAM46314	AAV4 VP3 coat prot
29	1690.5	52.0	544	24 AAG73941	Adeno-associated v
30	1676	51.6	534	16 AAR85386	Barbary duck parvo
31	1665	51.2	532	21 AAY58162	Adeno associated v
32	1665	51.2	532	23 AAU11407	Adeno-associated v
33	772	23.7	756	21 AAY71231	Capsid protein enc
34	497.5	15.3	781	16 AAM08986	Human parvovirus V
35	497.5	15.3	781	20 AAY23227	Erythrovirus V9 VP
36	493	15.2	781	24 ABP57263	Human parvovirus B
37	490	15.1	554	16 AAM08987	Human parvovirus V
38	490	15.1	554	20 AAY23230	Erythrovirus V9 VP
39	486	14.9	554	24 ABP57264	Human parvovirus B
40	486	14.9	554	24 ABP57267	Human parvovirus B
41	484	14.9	781	24 ABP57266	Human parvovirus B
42	476.5	14.7	543	12 AAR13405	Parvo virus B19 VP
43	310	9.5	264	12 AAR13407	Parvo virus B19 PA
44	295.5	9.1	370	12 AAR13406	Parvo virus B19 PA
45	283	8.7	202	23 AAU83481	Adeno-associated v

ALIGNMENTS

RESULT 1  
AAY711168  
ID AAY711168 standard; Protein; 599 AA.  
XX  
AC AAY711168;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 capsid protein VP2.  
XX  
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP2.  
XX  
OS Adeno associated virus serotype 1.  
XX  
PN WO200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
DR WPI; 2000-376571/32.  
XX  
DR N-PSDB; AAD00772, AAD00778.  
XX  
PT Novel adeno-associated virus serotype 1 polynucleotide useful for



PT preparation of medicament for delivery of a transgene to a host -

PS Claim 7; Page 93-95; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP2 which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 599 AA;

Query Match 100.0%; Score 3251; DB 21; Length 599;  
Best Local Similarity 100.0%; Pred. No. 1.5e-254;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAPGKKRPVEQSPQDPDSSSGIGKTGQPAKRLNFGQTGDSVPDPQLGEPATPAA 60  
DB 1 TAPGKKRPVEQSPQDPDSSSGIGKTGQPAKRLNFGQTGDSVPDPQLGEPATPAA 60  
QY 61 VGPTTMASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTWALPTNNHLY 120  
DB 61 VGPTTMASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTWALPTNNHLY 120  
QY 121 KOISSASTGASNDNHYFGYSTPWGYFDNRFCHFSPRDQRLINNNWGFPRKRLNFKLF 180  
DB 121 KOISSASTGASNDNHYFGYSTPWGYFDNRFCHFSPRDQRLINNNWGFPRKRLNFKLF 180  
QY 181 NIQKVEVTTNDGVTTIANNLSTVQVFSDSYQLPYVLGSAHQGLPPFPADVFMIPQYG 240  
DB 181 NIQKVEVTTNDGVTTIANNLSTVQVFSDSYQLPYVLGSAHQGLPPFPADVFMIPQYG 240  
QY 241 YLTLLNGSQAVGRSFFCYCLEYFPQMLRTGNNFTSYTFEVPFHSSVAHSQSLDRLMNP 300  
DB 241 YLTLLNGSQAVGRSFFCYCLEYFPQMLRTGNNFTSYTFEVPFHSSVAHSQSLDRLMNP 300  
QY 301 LIDQYLYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
DB 301 LIDQYLYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMGSGVMI FGKESAGASNTALDN 420  
DB 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMGSGVMI FGKESAGASNTALDN 420  
QY 421 VMITDEEIKATNPVATERFGTVANVFQSSSTDPATGDVHMGALPGMWQDRDYVLOGP 480  
DB 421 VMITDEEIKATNPVATERFGTVANVFQSSSTDPATGDVHMGALPGMWQDRDYVLOGP 480  
QY 481 IWAKIPIHTDGHFHPSPLMGGFGLKNPPPOILLKNTVPANPAEFASATKFAFITYSTG 540  
DB 481 IWAKIPIHTDGHFHPSPLMGGFGLKNPPPOILLKNTVPANPAEFASATKFAFITYSTG 540  
QY 541 QVSVEIWELOKQNSKRWNPVQVTSNYAKSANVDTVDNNGLYTEPRPIGTRYTRPL 599  
DB 541 QVSVEIWELOKQNSKRWNPVQVTSNYAKSANVDTVDNNGLYTEPRPIGTRYTRPL 599

RESULT 2

AAV711167

ID AAV711167 standard; Protein; 736 AA.

XX AAV711167;

AC AAV711167;

XX 08-SEP-2000 (first entry)

XX Adeno-associated virus serotype 1 capsid protein VP1.  
XX Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
XX cap protein; recombinant viral vector; gene delivery; gene therapy;  
XX vaccine; transgene; VP1.  
XX Adeno associated virus serotype 1.  
XX WO2000028061-A2.  
XX 18-MAY-2000.  
XX 02-NOV-1999; 99WO-US25694.  
XX 05-NOV-1998; 98US-0107114.  
XX (UYPE-) UNIV PENNSYLVANIA.  
XX Wilson JM, Xiao W;  
XX WPI; 2000-376571/32.  
XX N-PSDB; AAD00772, AAD00777.  
XX Novel adeno-associated virus serotype 1 polynucleotide useful for  
XX preparation of medicament for delivery of a transgene to a host -  
XX Claim 7; Page 87-90; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP1 which is  
CC useful in the production of recombinant viral vector for gene delivery.

SQ Sequence 736 AA;

Query Match 100.0%; Score 3251; DB 21; Length 736;  
Best Local Similarity 100.0%; Pred. No. 2e-254;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAPGKKRPVEQSPQDPDSSSGIGKTGQPAKRLNFGQTGDSVPDPQLGEPATPAA 60  
DB 138 TAPGKKRPVEQSPQDPDSSSGIGKTGQPAKRLNFGQTGDSVPDPQLGEPATPAA 197  
QY 61 VGPTTMASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTWALPTNNHLY 120  
DB 198 VGPTTMASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTWALPTNNHLY 257  
QY 121 KOISSASTGASNDNHYFGYSTPWGYFDNRFCHFSPRDQRLINNNWGFPRKRLNFKLF 180  
DB 258 KOISSASTGASNDNHYFGYSTPWGYFDNRFCHFSPRDQRLINNNWGFPRKRLNFKLF 317  
QY 181 NIQKVEVTTNDGVTTIANNLSTVQVFSDSYQLPYVLGSAHQGLPPFPADVFMIPQYG 240  
DB 318 NIQKVEVTTNDGVTTIANNLSTVQVFSDSYQLPYVLGSAHQGLPPFPADVFMIPQYG 377  
QY 241 YLTLLNGSQAVGRSFFCYCLEYFPQMLRTGNNFTSYTFEVPFHSSVAHSQSLDRLMNP 300  
DB 378 YLTLLNGSQAVGRSFFCYCLEYFPQMLRTGNNFTSYTFEVPFHSSVAHSQSLDRLMNP 437  
QY 301 LIDQYLYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
DB 438 LIDQYLYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 497

QY 361 NSNFTWTGASKYNLNGRESIIINPGTAMASHKODEDKFFPMGVMIFGKESAGASNTALDN 420  
 DB 498 NSNFTWTGASKYNLNGRESIIINPGTAMASHKODEDKFFPMGVMIFGKESAGASNTALDN 557  
 QY 421 VMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
 DB 558 VMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 617  
 QY 481 IWAKIPHTDGHFHPSPLMGFGGLKNPPPPQILIKNTVPANPPAEFSATKFAFPIQYSTG 540  
 DB 618 IWAKIPHTDGHFHPSPLMGFGGLKNPPPPQILIKNTVPANPPAEFSATKFAFPIQYSTG 677  
 QY 541 QVSVEIEWELOKENS KRWNPVEQVYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 599  
 DB 678 QVSVEIEWELOKENS KRWNPVEQVYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 736

RESULT 3  
 AAB59847  
 ID AAB59847 standard; Protein; 736 AA.  
 XX  
 AC AAB59847;  
 XX  
 AC  
 DT 28-MAR-2001 (first entry)  
 XX  
 DE AAV6 capsid protein VP1.  
 XX  
 KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
 KW atherosclerosis; sickle cell anaemia; thalassaemia;  
 KW blood clotting disorder; diabetes; capsid protein VP1.  
 XX  
 OS Adeno associated virus.  
 XX  
 XX US6156303-A.  
 PN  
 PD 05-DEC-2000.  
 XX  
 PF 11-JUN-1997; 97US-0873168.  
 XX  
 PR 11-JUN-1997; 97US-0873168.  
 XX  
 PA (UNIW ) UNIV WASHINGTON.  
 XX  
 PI Russell DW, Rutledge EA;  
 XX  
 DR WPI; 2001-060164/07.  
 XX  
 PT Adeno-associated virus serotype 6 and viral vector derived from it for  
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
 PT syndrome, sickle cell anemia, thalassaemia and diabetes -  
 XX  
 PS Disclosure; Fig 2; 50pp; English.  
 XX  
 CC The present invention relates to adeno-associated virus serotypes. The  
 CC present sequence is capsid protein VP1 of one such serotype (AAV6). AAV6  
 CC can be used to construct AAV viral vectors for use in gene therapy for a  
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
 CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
 CC The AAV viral vectors have increased transduction efficiency of a  
 CC particular host cell as the AAV virion containing the AAV vector genome  
 CC can be modified to express a capsid protein of an AAV serotype that  
 CC transduces the selected host cell.  
 XX  
 SQ Sequence 736 AA;

Query Match 99.3%; Score 3229; DB 22; Length 736;  
 Best Local Similarity 99.2%; Pred. No. 1.2e-252;  
 Matches 594; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TAPGKKRPVEQSPQEPDSSGIGKTQQPAKKRLNFGQTGDSVPDPQPLGEPATPAA 60  
 DB 138 TAPGKKRPVEQSPQEPDSSGIGKTQQPAKKRLNFGQTGDSVPDPQPLGEPATPAA 197

QY 61 VGPTTMASSGGAPMADNNEGADGVGNASGNHCHDSTWLGDRTVITSTRTALPTYNHLY 120  
 DB 198 VGPTTMASSGGAPMADNNEGADGVGNASGNHCHDSTWLGDRTVITSTRTALPTYNHLY 257  
 QY 121 KOISSASTGASNDNHFGYSTPMWGYDFNRFHCHFSPRDWQRLINNNMGFRPKRLNFKLF 180  
 DB 258 KOISSASTGASNDNHFGYSTPMWGYDFNRFHCHFSPRDWQRLINNNMGFRPKRLNFKLF 317  
 QY 181 NIQVKEVTNDGVTTIANNLSTVQVFSDESEYQLPVVLGSAHQGLPPPADVFMIPQVG 240  
 DB 318 NIQVKEVTNDGVTTIANNLSTVQVFSDESEYQLPVVLGSAHQGLPPPADVFMIPQVG 377  
 QY 241 YLTLANGSAQVGRSSFYCLEYPPSQMLRTGNNTFTSYTFEYPPHSSVAHSQSLDRLMNP 300  
 DB 378 YLTLANGSAQVGRSSFYCLEYPPSQMLRTGNNTFTSYTFEYPPHSSVAHSQSLDRLMNP 437  
 QY 301 LIDQYLYLNRNTQNGSQAQNKDLLFSRGSAGMSVQPKNWLPGCYQQRVSKTKTDNN 360  
 DB 438 LIDQYLYLNRNTQNGSQAQNKDLLFSRGSAGMSVQPKNWLPGCYQQRVSKTKTDNN 497  
 QY 361 NSNFTWTGASKYNLNGRESIIINPGTAMASHKODEDKFFPMGVMIFGKESAGASNTALDN 420  
 DB 498 NSNFTWTGASKYNLNGRESIIINPGTAMASHKODEDKFFPMGVMIFGKESAGASNTALDN 557  
 QY 421 VMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
 DB 558 VMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 617  
 QY 481 IWAKIPHTDGHFHPSPLMGFGGLKNPPPPQILIKNTVPANPPAEFSATKFAFPIQYSTG 540  
 DB 618 IWAKIPHTDGHFHPSPLMGFGGLKNPPPPQILIKNTVPANPPAEFSATKFAFPIQYSTG 677  
 QY 541 QVSVEIEWELOKENS KRWNPVEQVYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 599  
 DB 678 QVSVEIEWELOKENS KRWNPVEQVYTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 736

## RESULT 4

AAV71169  
 ID AAV71169 standard; Protein; 534 AA.  
 XX  
 AC AAV71169;  
 XX  
 DT 08-SEP-2000 (first entry)  
 XX  
 DE Adeno-associated virus serotype 1 capsid protein VP3.  
 XX  
 KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
 KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
 KW vaccine; transgene; VP3.  
 XX  
 OS Adeno associated virus serotype 1.  
 XX  
 PN WO200028061-A2.  
 XX  
 PD 18-MAY-2000.  
 XX  
 PF 02-NOV-1999; 99WO-US25694.  
 XX  
 PR 05-NOV-1998; 98US-0107114.  
 XX  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 XX  
 PI Wilson JM, Xiao W;  
 XX  
 DR WPI; 2000-376571/32.  
 DR N-P5DB; AAD00772, AAD00779.  
 XX  
 PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
 PT preparation of medicament for delivery of a transgene to a host -  
 XX  
 PS Claim 7; Page 99-101; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP3 which is  
CC useful in the production of recombinant viral vector for gene delivery.  
XX  
SQ Sequence 534 AA;  
Query Match 89.4%; Score 2906; DB 21; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.1e-226;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 66 MASGGAPMADNNEGADGVNASGNHCHDSTWLGDRVITTTTRTWALPTNNHLYKQIS 125  
DB 1 MASGGAPMADNNEGADGVNASGNHCHDSTWLGDRVITTTTRTWALPTNNHLYKQIS 60  
QY 126 ASTGASNDNHFGYSTPMWGYDFNRRFCHFSPRDWORLNNNWRPRLNFKLNIQVK 185  
DB 61 ASTGASNDNHFGYSTPMWGYDFNRRFCHFSPRDWORLNNNWRPRLNFKLNIQVK 120  
QY 186 EVTTNDGVTTIANLLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOYGYLTN 245  
DB 121 EVTTNDGVTTIANLLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOYGYLTN 180  
QY 246 NGSQAVGRSSFYCLEYPPSQMLRTGNFTSYTFEEVPHSSVAHSQSLDRLNPLIDQY 305  
DB 181 NGSQAVGRSSFYCLEYPPSQMLRTGNFTSYTFEEVPHSSVAHSQSLDRLNPLIDQY 240  
QY 306 LYTLLNRTONSGSAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQQRVSKTKTDNNNSFT 365  
DB 241 LYTLLNRTONSGSAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQQRVSKTKTDNNNSFT 300  
QY 366 WTGASKYNLNGRESIINPGTAMASHKDDEKFFPMGSGVMI FGKESAGASNTALDNYMTD 425  
DB 301 WTGASKYNLNGRESIINPGTAMASHKDDEKFFPMGSGVMI FGKESAGASNTALDNYMTD 360  
QY 426 EBEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGPWAKI 485  
DB 361 EBEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGPWAKI 420  
QY 486 PHTDGHFHPSPLMGGFGLKNPPQIILIKNTVPANPPAEFSATKPFASFTOYSTGVSV 545  
DB 421 PHTDGHFHPSPLMGGFGLKNPPQIILIKNTVPANPPAEFSATKPFASFTOYSTGVSV 480  
QY 546 IEWELOKENSKRNPEVQYTSNVAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 599  
DB 481 IEWELOKENSKRNPEVQYTSNVAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 534  
RESULT 5  
AAB59846  
ID AAB59846 standard; Protein; 736 AA.  
XX  
AC AAB59846;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE AAV3B capsid protein VP1.  
XX  
KW AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KW atherosclerosis; sickle cell anaemia; thalassemia;  
XX blood clotting disorder; diabetes; capsid protein VP1.  
XX

OS Adeno associated virus.  
PN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
DR WPI; 2001-060164/07.  
XX  
PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes -  
XX  
PS Disclosure; Fig 2; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV3B).  
CC AAV3B can be used to construct AAV viral vectors for use in gene therapy  
CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 736 AA;  
Query Match 87.1%; Score 2832; DB 22; Length 736;  
Best Local Similarity 85.7%; Pred. No. 1.8e-220;  
Matches 514; Conservative 34; Mismatches 50; Indels 2; Gaps 2;  
QY 1 TAPGKKRPVEGSPQPSDSSSGTGKQOPAKRLNFGTGDSESVDPDQPLGEPPATPAA 60  
DB 138 TAPGKKRPVDSQPSDSSSGTGKQOPAKRLNFGTGDSESVDPDQPLGEPPATPAA 197  
QY 61 VQPTTASGGGAPMADNNEGADGVNASGNHCHDSTWLGDRVITTTTRTWALPTNNHLY 120  
DB 198 LGSNTMASGGGAPMADNNEGADGVNASGNHCHDSTWLGDRVITTTTRTWALPTNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPMWGYDFNRRFCHFSPRDWORLNNNWRPRLNFKL 180  
DB 258 KOISSQS-GASNDNHFGYSTPMWGYDFNRRFCHFSPRDWORLNNNWRPRLNFKL 316  
QY 181 NIQVKEVTNDGVTTIANLLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOY 240  
DB 317 NIQVKEVTNDGVTTIANLLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOY 376  
QY 241 YLTLLNGSAVGRSSFYCLEYPPSQMLRTGNFTSYTFEEVPHSSVAHSQSLDRLNMP 300  
DB 377 YLTLLNGSAVGRSSFYCLEYPPSQMLRTGNFTSYTFEEVPHSSVAHSQSLDRLNMP 436  
QY 301 LIDQYLYLNRTO-NQSGAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQQRVSKTKTD 359  
DB 437 LIDQYLYLNRTO-NQSGAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQQRVSKTKTD 496  
QY 360 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDEKFFPMGSGVMI FGKESAGASNTALD 419  
DB 497 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDEKFFPMGSGVMI FGKESAGASNTALD 556  
QY 420 NVMITDDEBEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQ 479  
DB 557 NVMITDDEBEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQ 616  
QY 480 PIWAKIPHTDGHFHPSPLMGGFGLKNPPQIILIKNTVPANPPAEFSATKPFASFTOYST 539  
DB 617 PIWAKIPHTDGHFHPSPLMGGFGLKNPPQIILIKNTVPANPPAEFSATKPFASFTOYST 676

QY 540 GOVSVEIWELOKENS KRNWPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
 DB 677 GOVSVEIWELOKENS KRNWPEIQYTSNYKSNVNDFTVDNNGVYSEPRPIGTRYLTRNL 736

RESULT 6  
 AAB59845  
 ID AAB59845 standard; Protein; 736 AA.  
 XX AAB59845;  
 AC  
 DT 28-MAR-2001 (first entry)  
 XX AAV3A capsid protein VP1.  
 DE  
 DE  
 KW AAV3A; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
 KW atherosclerosis; sickle cell anaemia; thalassaemia;  
 KW blood clotting disorder; diabetes; capsid protein VP1.  
 XX Adeno associated virus.  
 OS  
 XX US6156303-A.  
 XX  
 XX 05-DEC-2000.  
 XX  
 XX 11-JUN-1997; 97US-0873168.  
 XX  
 XX 11-JUN-1997; 97US-0873168.  
 XX  
 XX (UNIW ) UNIV WASHINGTON.  
 XX  
 XX Russell DW, Rutledge EA;  
 XX WPI; 2001-060164/07.  
 XX  
 XX Adeno-associated virus serotype 6 and viral vector derived from it for  
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
 PT syndrome, sickle cell anemia, thalassaemia and diabetes  
 XX  
 XX Disclosure; Fig 2; 50pp; English.  
 XX  
 XX The present invention relates to adeno-associated virus serotypes. The  
 CC present sequence is capsid protein VP1 of one such serotype (AAV3A).  
 CC AAV3A can be used to construct AAV viral vectors for use in gene therapy  
 CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
 CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
 CC The AAV viral vectors have increased transduction efficiency of a  
 CC particular host cell as the AAV virion containing the AAV vector genome  
 CC can be modified to express a capsid protein of an AAV serotype that  
 CC transduces the selected host cell.  
 XX  
 XX Sequence 736 AA;  
 SQ

Query Match 86.6%; Score 2815; DB 22; Length 736;  
 Best Local Similarity 85.2%; Pred. No. 4.2e-219;  
 Matches 511; Conservative 35; Mismatches 52; Indels 2; Gaps 2;

QY 1 TAPGKKRPVQSPOEPDSSGIGTGOQPAKRLNFGOTGDSVDPDPQLGPRPATPA 60  
 DB 138 TAPGKGVQSQEPDPSSSGVSGKQPAKRLNFGOTGDSVDPDPQLGPRPATPS 197  
 QY 61 VGPTTMAAGCGAPMADNNEGADGVGNAGNHCDSWTWLGDRVITTTSTRTWALPTNNHLY 120  
 DB 198 LGSNTMASGCGAPMADNNEGADGVGNSSGNWCHDSQWLGDRVITTTSTRTWALPTNNHLY 257  
 QY 121 KQISSASTGASNDNHFGYSTPWCYDFNPFCHFSRDRNQRLLNNWGRPRKLNPKLF 190  
 DB 258 KQISSQS-GASNDNHFGYSTPWCYDFNPFCHFSRDRNQRLLNNWGRPRKLNPKLF 316  
 QY 181 NIOKVEYTTNDGVTIANNTLSTVQVPSDSEYQLPYVLGSAHQGCLPPFPFADVFMIPOYG 240  
 DB 317 NIOQVGTQNDGVTIANNTLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPFADVFMIPOYG 376

QY 241 YLTNNGSQAVGRSSFCLEYFPSSQMLRTGNNFTSYTPEEVPFHSYAHQSOLDRLMNP 300  
 DB 377 YLTNNGSQAVGRSSFCLEYFPSSQMLRTGNNFQFSYTTEDVPFHSYAHQSOLDRLMNP 436  
 QY 301 LIDQLYLYLNRTQ-NQSGSAQNKDLLFSRGSAGMSVQPKNMLPGPCYRQORVSKTKTON 359  
 DB 437 LIDQLYLYLNRTQGTTSNTQSRLLFSAQPSQMSLOARNMLPGPCYRQORLSKTANDN 496  
 QY 360 NNSNFTWTCASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIPGKESAGASNTALD 419  
 DB 497 NNSNFTWTAASKYHLNGLRSLVNPGPAMASHKDDKDFPMSGVMIPGKESAGASNTALD 556  
 QY 420 NYMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQ 479  
 DB 557 NYMITDEEIRTNPVATEQYGTVANLQSSNTAPTGTGVNHQALPGMWQDRDVLQ 616  
 QY 480 PIWAKIPHTDGHFSPGLMGFGGLKNPPQIILIKNTVPANPPAEPASATKPAFIQYST 539  
 DB 617 PIWAKIPHTDGHFSPGLMGFGGLKHPPIQIMIKNTVPANPPPTTSPAKFASFIQYST 676  
 QY 540 GOVSVEIWELOKENS KRNWPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
 DB 677 GOVSVEIWELOKENS KRNWPEIQYTSNYKSNVNDFTVDNNGVYSEPRPIGTRYLTRNL 736

RESULT 7  
 AAG65792  
 ID AAG65792 standard; Protein; 735 AA.  
 XX AAG65792;  
 AC  
 DT 11-FEB-2002 (first entry)  
 XX  
 XX Adeno-associated virus 2 (AAV-2) major coat protein VP1.  
 DE  
 DE Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
 KW inverted terminal repeat; neurotropic; neuroprotective; antianemic; ITR;  
 KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
 KW major coat protein; AAV-2; VP1.  
 XX  
 XX Adeno-associated virus 2.  
 OS  
 XX WO200168888-A2.  
 XX  
 XX 20-SEP-2001.  
 XX  
 XX 13-MAR-2001; 2001WO-US07927.  
 XX  
 XX 14-MAR-2000; 2000US-189110P.  
 XX  
 XX (NEUR-) NEUROLOGIX INC.  
 XX  
 XX Xiao W, During MJ;  
 XX WPI; 2001-596912/67.  
 XX N-PSDB; AAI66974.  
 XX  
 XX Recombinant viral vector useful in improving gene therapy in a subject,  
 PT and for increasing efficiency of entry into a cell, comprises a  
 PT chimeric capsid having one non-native amino acid sequence and a desired  
 PT transgene  
 XX  
 XX Disclosure; Page 50; 53pp; English.  
 XX  
 XX The invention provides a recombinant viral vector (RVV) comprising a  
 CC chimeric capsid (I) having at least one non-native amino acid sequence,  
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
 CC or their combination, and a transgene flanked 5' and 3' by inverted  
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
 CC combination. The RVV is useful for improving gene therapy in a subject  
 CC with a disorder, and for increasing the efficiency of entry into a cell,  
 CC which involves producing (I) encapsulating a viral vector, and contacting



Db 677 QVSVEIEWELQKNSKRWNPFIQYTSNYSKNSVNVDFVTDTNGVYSBPRPIGTRYLTRNL 735

RESULT 9  
AAB59844  
ID AAB59844 standard; Protein; 735 AA.  
XX AAB59844;  
AC  
DT 28-MAR-2001 (first entry)  
XX AAV2 capsid protein VP1.  
DE  
XX AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KW atherosclerosis; sickle cell anemia; thalassemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
XX  
OS Adeno associated virus.  
XX  
PN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
XX WPI; 2001-060164/07.  
DR  
XX  
XX  
PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX  
XX  
PS Claim 7; Fig 2; 50pp; English.  
XX  
XX The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV2). AAV2  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anemia, thalassemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 735 AA;

Query Match 85.0%; Score 2764.5; DB 22; Length 735;  
Best Local Similarity 82.8%; Pred. No. 5.1e-215;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKKRPVEHSQVPEPSSSGIGTQQQAKXKLNFGQTGDSVDPQPLGPPATPAA 60  
Db 138 TAPGKKRPVEHSQVPEPSSSGTCKAGQQAPARKLNFGQTGDSVDPQPLGPPAPSG 197  
QY 61 VGPPTWASGGAPKADNNEGADGVNAGNWHCDSTWLGDRVITTTTRTWALPTYNHLY 120  
Db 198 LGNTWATGSGAPKADNNEGADGVNAGNWHCDSTWMDGRVITTTTRTWALPTYNHLY 257  
QY 121 KOISSASTGASNDNHVFGYSTPGYDFNRFHCFSPRDWQRLNNWGFPRKRLNFKLF 180  
Db 258 KQISSQS-GASNDNHVFGYSTPGYDFNRFHCFSPRDWQRLNNWGFPRKRLNFKLF 316  
QY 181 NIQVKEVTTNDGVTTIANNLTSTVQVPSDSYQLPYVLGSAHQGCLPPFPADVFMIPQYG 240  
Db 317 NIQVKEVTTNDGVTTIANNLTSTVQVPSDSYQLPYVLGSAHQGCLPPFPADVFMIPQYG 376  
QY 241 YLTLLNGSQAVGRSSFCLEYPPSQMLRTGNFTTSYTFEDVPFHSSYAHSSQSLDLRMLNP 300

Db 377 YLTLLNGSQAVGRSSFCLEYPPSQMLRTGNFTTSYTFEDVPFHSSYAHSSQSLDLRMLNP 436  
QY 301 LIDQVLYLNRNTQNGSGSAQNKDLFLSRGSPAGMSVQPKWMLPGPCYRQORYSKTKTDNN 360  
Db 437 LIDQVLYLNRNTQNGSGSAQNKDLFLSRGSPAGMSVQPKWMLPGPCYRQORYSKTKTDNN 496  
QY 361 NSNFTWTGASKYNLNGRBSIINPGTAMASHKDDDEKFFPMGSGVMIFGKESAGASNTALDN 420  
Db 497 NSEYSWTGATKYHLNGRDSLVNPGPAMASHKDDDEKFFPMGSGVMIFGKESAGASNTALDN 556  
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHANGALPGMWQDRDYYLQGP 480  
Db 557 VMITDEEIRTTNPVATEQYGSVSTNLRGNRQAATADVNTQGVLPGMVWQDRDYYLQGP 616  
QY 481 IWAKIPHTDGHFHPSPMLMGGLKMPPOILLKNTVPANPPAEFSATKPFASFITOYSTG 540  
Db 617 IWAKIPHTDGHFHPSPMLMGGLKMPPOILLKNTVPANPPAEFSATKPFASFITOYSTG 676  
QY 541 QVSVEIEWELQKNSKRWNPFIQYTSNYSKNSVNVDFVTDTNGVYSBPRPIGTRYLTRNL 599  
Db 677 QVSVEIEWELQKNSKRWNPFIQYTSNYSKNSVNVDFVTDTNGVYSBPRPIGTRYLTRNL 735

RESULT 10  
AAU98974  
ID AAU98974 standard; Protein; 735 AA.  
XX AAU98974;  
AC  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Adeno-associated virus 2 (AAV2) vector, VP1 capsid protein.  
KW Adeno-associated virus 2 vector; AAV2; cancer; VP1 capsid;  
KW heparin-sulphate proteoglycan; vaccine; immune response;  
KW ovarian cancer.  
XX  
OS Adeno-associated virus 2.  
XX  
PN WO200253703-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 04-JAN-2002; 2002WO-US00152.  
XX  
PR 05-JAN-2001; 2001US-260124P.  
XX  
PA (CHIL-) CHILDRENS HOSPITAL INC.  
XX  
PI Bartlett JS;  
XX  
DR WPI; 2002-583608/62.  
XX  
XX N-PSDB; ABK89694.  
PT New adeno-associated virus vector comprises a biotinylated capsid or  
PT capsid protein with an amino acid insertion in the VP1 capsid, useful  
PT as a vaccine or for transferring a therapeutic peptide to a cancer cell  
XX  
XX  
PS Claim 1; Page 49-51; 57pp; English.  
XX  
XX The invention relates to an adeno-associated virus (AAV) vector (I)  
CC comprising a biotinylated capsid or capsid protein (II) with an amino  
CC acid insertion following the capsid amino acid at position 139, 161, 588  
CC or 657 in the VP1 capsid. The AAV vector comprises a capsid protein  
CC containing one or more amino acid insertions that ablate the ability of  
CC the vector to bind heparin-sulphate proteoglycan and allow the vector to  
CC use a cellular receptor not used by wild type AAV. Modified (I) are  
CC useful as vaccines to elicit immune responses to amino acids, where the  
CC response can be protective and/or therapeutic. (I) may be used to  
CC transfer a therapeutic peptide to a cancer cell, particularly to an  
CC ovarian cancer cell. The present sequence represents the adeno-associated

```
CC virus 2 (AAV2) vector, VP1 capsid protein used in to make modified
CC AAV2 vectors.
XX
SQ Sequence 735 AA;

Query Match      85.0%; Score 2764.5; DB 23; Length 735;
Best Local Similarity 82.8%; Pred. No. 5.1e-215;
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1  TAPGKKRPVEQSPOEPDSSSGIGKTCQPAKRLNFGQTDSEVPDPQPLGEPATPA 60
DB 138 TAPGKKRPVEHSPEVPDSSSGTGKAGQPAKRLNFGQTDADSVDPDQPLGQPPA 197
QY 61  VGTPTWASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTRTWALPTNNHLY 120
DB 198 LGTNTWATGSGAPMADNNEGADGVGNAGNWHCDSTWMDGRVITTTTRTWALPTNNHLY 257
QY 121 KQISSASTGASNDNHVFGYSTPMGYDFNRFCHFSRDMQRLINNNWGFPRKRLNFKLP 180
DB 258 KQISSQS-GASNDNHVFGYSTPMGYDFNRFCHFSRDMQRLINNNWGFPRKRLNFKLP 316
QY 181 NIOQKEVTNDGVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGLPPFPADVFMIPOYG 240
DB 317 NIOQKEVTQNDGVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGLPPFPADVFMIPOYG 376
QY 241 YLTNNGSOAVGRSSPYCLEYFPSPQMLRTGNNTFFSYTPEVFFHSSYAHQSGLDLNMP 300
DB 377 YLTNNGSOAVGRSSPYCLEYFPSPQMLRTGNNTFFSYTPEVFFHSSYAHQSGLDLNMP 436
QY 301 LIDQYLYLNRNTQNSQAGNKKDLFSRSGPAGMSVQPKNLPGPCYRQORVSKTKTDNN 360
DB 437 LIDQYLYLNRTPSGTTSQSLQFSQAGSDIRDSQRLNPGPCYRQORVSKTSADNN 496
QY 361 NSNFTWTGASKYNLNGRESITINPGTAWASHKDEKDFPMGVMIFGKSAGASNTALDN 420
DB 497 NSBYSTWGTAKYHLNGRDSLVNPGPAMASHKDEKDFPMGVMIFGKSGERTNVDIEK 556
QY 421 VMTDEEIKATNPVATEREFTVAVNPFQSSSTDPATGCVHAMGALPGVMWQDRDVLQGP 480
DB 557 VMTDEEIRTPVATEQYGVSTNLQGRNQAAATADVNTQGLPGVMWQDRDVLQGP 616
QY 481 IWAKIPHTDGHFPLMGFGGLNKPQPLIKNTVPANPPAPFASFIQYSTG 540
DB 617 IWAKIPHTDGHFPLMGFGGLNKPQPLIKNTVPANPFSTTFAAKFASFIQYSTG 676
QY 541 QVSVEIWELOKENSKEWNPVOYTSNYAKSANVDFTVDNNGLYTSRPTGTRYLTP 599
DB 677 QVSVEIWELOKENSKEWNPVOYTSNYAKSANVDFTVDNNGLYTSRPTGTRYLTP 735

RESULT 11
AAG65793
ID AAG65793 standard; Protein; 598 AA.
XX
AC AAG65793;
XX
DT 11-FEB-2002 (first entry)
XX
DE Adeno-associated virus 2 (AAV-2) major coat protein VP2.
XX
KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;
KW inverted terminal repeat; nontropic; neuroprotective; antianemic; ITR;
KW antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;
KW major coat protein; AAV-2; VP2.
XX
OS Adeno-associated virus 2.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "encoded by ACG"
XX
XX WO200168888-A2.
XX
```

```
PD 20-SEP-2001.
XX
PF 13-MAR-2001; 2001WO-US07927.
XX
PR 14-MAR-2000; 2000US-189110P.
PA (NEUR-) NEUROLOGIX INC.
XX
PI Xiao W, During MJ;
XX
DR WPI; 2001-596912/67.
XX
DR N-PSDB; AAI66974.
XX
PT Recombinant viral vector useful in improving gene therapy in a subject,
PT and for increasing efficiency of entry into a cell, comprises a
PT chimeric capsid having one non-native amino acid sequence and a desired
PT transgene
PS
PS Disclosure; Page 51; 53pp; English.
XX
CC The invention provides a recombinant viral vector (RVV) comprising a
CC chimeric capsid (I) having at least one non-native amino acid sequence,
CC derived from a capsid protein domain of parvovirus (II), a virus (III),
CC or their combination, and a transgene flanked 5' and 3' by inverted
CC terminal repeat (ITR) sequences, derived from (II), (III), or their
CC combination. The RVV is useful for improving gene therapy in a subject
CC with a disorder, and for increasing the efficiency of entry into a cell,
CC which involves producing (I) encapsulating a viral vector, and contacting
CC a cell with RVV having (I) such that (I) binds to an attachment site on
CC the cell surface and permits the vector to enter the cell efficiently.
CC A pharmaceutical composition comprising RVV with (I) containing a
CC transgene sequence associated with a disease or a disorder such that
CC expression of the transgene would result in amelioration of the disease
CC or disorder such as inherited neurological and metabolic diseases e.g.
CC lysosomal storage disease, Leech-Nyhan syndrome, amyloid polyneuropathy,
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
CC fibrosis, diabetes, diseases associated with hormone deficiencies,
CC retinoblastoma and various types of neoplastic cells which include
CC tumors especially central nervous system tumors, neoplasms, carcinomas,
CC sarcomas, leukemias and lymphoma. The present sequence represents the
CC adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences
CC are used in the construction of a chimeric vector.
XX
SQ Sequence 598 AA;

Query Match      84.9%; Score 2759.5; DB 22; Length 598;
Best Local Similarity 82.8%; Pred. No. 9.7e-215;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2  APGKKRPVEQSPOEPDSSSGIGKTCQPAKRLNFGQTDSEVPDPQPLGEPATPA 61
DB 2  APGKKRPVEHSPEVPDSSSGTGKAGQPAKRLNFGQTDADSVDPDQPLGQPPA 61
QY 62  GPTTWASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTRTWALPTNNHLY 121
DB 62  GTNTWATGSGAPMADNNEGADGVGNAGNWHCDSTWMDGRVITTTTRTWALPTNNHLY 121
QY 122 QISSASTGASNDNHVFGYSTPMGYDFNRFCHFSRDMQRLINNNWGFPRKRLNFKLP 181
DB 122 QISSQS-GASNDNHVFGYSTPMGYDFNRFCHFSRDMQRLINNNWGFPRKRLNFKLP 180
QY 182 IQQKEVTNDGVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGLPPFPADVFMIPOYG 241
DB 181 IQQKEVTQNDGVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGLPPFPADVFMIPOYG 240
QY 242 LTLNNGSOAVGRSSPYCLEYFPSPQMLRTGNNTFFSYTPEVFFHSSYAHQSGLDLNMP 301
DB 241 LTLNNGSOAVGRSSPYCLEYFPSPQMLRTGNNTFFSYTPEVFFHSSYAHQSGLDLNMP 300
QY 302 IDQYLYLNRNTQNSQAGNKKDLFSRSGPAGMSVQPKNLPGPCYRQORVSKTKTDNN 361
DB 301 IDQYLYLNRTPSGTTSQSLQFSQAGSDIRDSQRLNPGPCYRQORVSKTSADNN 360
```





or 657 in the VP1 capsid. The AAV vector comprises a capsid protein containing one or more amino acid insertions that ablate the ability of the vector to bind heparin-sulphate proteoglycan and allow the vector to use a cellular receptor not used by wild type AAV. Modified (I) are the useful as vaccines to elicit immune responses to amino acids, where the response can be protective and/or therapeutic. (I) may be used to transfer a therapeutic peptide to a cancer cell, particularly to an ovarian cancer cell. The present sequence represents the adeno-associated virus 2 (AAV2) vector, VP2 capsid protein used in to make modified AAV2 vectors.

XX Sequence 598 AA;

Query Match 84.9%; Score 2759.5; DB 23; Length 598;  
Best Local Similarity 82.8%; Pred. No. 9.7e-215;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2 APGKKRPVQSPQEPDSSGIGTKTQQAQKRLNFGQTDSEVPDPQLEPPATPAAV 61  
DB 2 APGKKRPVHSPVEPDSSSGTGKAGQQAQKRLNFGQTDADSVDPQPLQPPAAPSG 61  
QY 62 GPTTASGCGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYK 121  
DB 62 GNTWTATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVTITSTRTWALPTNNHLYK 121  
QY 122 QISSASTGASNDNHYFGYSTPMGYDFNRFCHFSFRDQWRLNINNWGFRPKRLNFKLFN 181  
DB 122 QISSQS-GASNDNHYFGYSTPMGYDFNRFCHFSFRDQWRLNINNWGFRPKRLNFKLFN 180  
QY 182 IQVKEVTNDGVTITANNLTSTVQVPSDSEYQLPYVLGSAHQCLPPFPADVFMIPQYQ 241  
DB 181 IQVKEVTNDGVTITANNLTSTVQVPSDSEYQLPYVLGSAHQCLPPFPADVFMIPQYQ 240  
QY 242 LTLNNGSOAVGRSSFYCLEYFPSSQMLRTGNFTFSEYVPHSSYAHSSQSLDRMLNPL 301  
DB 241 LTLNNGSOAVGRSSFYCLEYFPSSQMLRTGNFTFSEYVPHSSYAHSSQSLDRMLNPL 300  
QY 302 IDQYLYLNRTQNGSQAQNKDLLFSRGSAGMSVQPKNWLPGPCYRQORVSKTKTDNNN 361  
DB 301 IDQYLYLSRTNTPSGTITQSLQFSQAGSDIRQSRNWLPGPCYRQORVSKTSADNN 360  
QY 362 SNFTWTGAKYNLNGRESIINPGTAMASHKDDKFFPMGSGVMI FKGESAGASNTALDNV 421  
DB 361 SEYSWTGATKYHLNGRDSLNVNPGPAMASHKDDKFFPMGSGVMI FKGESAGASNTALDNV 420  
QY 422 MITDEBEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 481  
DB 421 MITDEBEIRTNVPATEQVGSVSTNLQRNQAATADVNTQGVLPQMWQDRDVLQGP 480  
QY 482 WAKIPHTDGHFHPSPLMGGFGLKHPPPQILIKNTVPANPPAEFSAATKFAFITQYSTGQ 541  
DB 481 WAKIPHTDGHFHPSPLMGGFGLKHPPPQILIKNTVPANPPAEFSAATKFAFITQYSTGQ 540  
QY 542 VSVLEIWELOKENSXRNWPEVOYTSNYAKSANVDVTVDNNGLYTEPRPIGRYLRPL 599  
DB 541 VSVLEIWELOKENSXRNWPEIOYTSNYAKSANVDVTVDNNGLYTEPRPIGRYLRPL 598

RESULT 14

AAB50326

ID AAB50326 standard; protein; 734 AA.

XX

AC AAB50326;

XX 09-MAR-2001 (first entry)

DE Adeno-associated virus capsid protein sequence.

XX Adeno-associated virus; AAV; capsid; virus binding inhibition;

KW competitive inhibitor.

XX Mastadenovirus.

OS Synthetic.

XX WO200073316-A2.  
XX 07-DEC-2000.  
XX 26-MAY-2000; 2000WO-US14466.  
XX 28-MAY-1999; 99US-0321589.  
XX (CELL-) CELL GENESYS INC.  
XX Patel S, McArthur J;  
XX WPI; 2001-061507/07.  
XX New polypeptide portion of a virus protein binding to an antibody  
XX specific for the virus useful for inhibiting binding of the virus to a  
XX cell or for binding host antibody to provide a transient tolerant or  
XX non-responsive state -  
XX Disclosure; Fig 2; 33pp; English.

The present sequence is given in a specification relating to polypeptide portions of a virus protein or its derivative, that bind to an antibody specific for the virus or inhibit binding of the virus to a cell. The polypeptides are used in inhibiting the binding of viruses to cells of a host. Oligopeptides that inhibit binding of virus to receptor can be used as competitive inhibitors to release bound virus in an adsorption-type assay, and if an antibody was used as an immunoadsorbent, the oligopeptide could be used to elute bound virus from a solid support to which virus antibody is immobilised. These oligopeptides may further be used to bind to host antibody to provide a transient tolerant or non-responsive state.

SQ Sequence 734 AA;

Query Match 84.2%; Score 2738; DB 22; Length 734;  
Best Local Similarity 82.5%; Pred. No. 7.2e-213;  
Matches 494; Conservative 43; Mismatches 60; Indels 2; Gaps 2;

QY 1 TAPGKKRPVQSPQEPDSSGIGTKTQQAQKRLNFGQTDSEVPDPQLEPPATPA 60  
DB 138 TAPGKKRPVHSPVEPDSSSGTGKAGQQAQKRLNFGQTDADSVDPQPLQPPAAPSG 197  
QY 61 VGPTTASGCGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY 120  
DB 198 LGTWTATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVTITSTRTWALPTNNHLY 257  
QY 121 QISSASTGASNDNHYFGYSTPMGYDFNRFCHFSFRDQWRLNINNWGFRPKRLNFKLF 180  
DB 258 QISSQS-GASNDNHYFGYSTPMGYDFNRFCHFSFRDQWRLNINNWGFRPKRLNFKLF 316  
QY 181 NIQVKEVTNDGVTITANNLTSTVQVPSDSEYQLPYVLGSAHQCLPPFPADVFMIPQY 240  
DB 317 NIQVKEVTNDGVTITANNLTSTVQVPSDSEYQLPYVLGSAHQCLPPFPADVFMIPQY 376  
QY 241 YLTLLNNGSOAVGRSSFYCLEYFPSSQMLRTGNFTFSEYVPHSSYAHSSQSLDRMLN 300  
DB 377 YLTLLNNGSOAVGRSSFYCLEYFPSSQMLRTGNFTFSEYVPHSSYAHSSQSLDRMLN 436  
QY 301 LIDQYLYLNRTQNGSQAQNKDLLFSRGSAGMSVQPKNWLPGPCYRQORVSKTKTDNN 360  
DB 437 LIDQYLYLSRTNTPSGTITQSLQFSQAGSDIRQSRNWLPGPCYRQORVSKTSADNN 496  
QY 361 NSFTWTGAKYNLNGRESIINPGTAMASHKDDKFFPMGSGVMI FKGESAGASNTALDN 420  
DB 497 NSEYSWTGATKYHLNGRDSLNVNPGPAMASHKDDKFFPMGSGVMI FKGESAGASNTALDN 556  
QY 421 WMITDEBEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
DB 557 WMITDEBEIRTTN-VATEQVGSVSTNLQRNQAATADVNTQGVLPQMWQDRDVLQGP 615  
QY 481 IWAKIPHTDGHFHPSPLMGGFGLKHPPPQILIKNTVPANPPAEFSAATKFAFITQYSTG 540

Db 616 IWAKIPHTDGHFSPPLMGFGGLKHPPPQILIKNTVPANPSTTFSAAKFASFIQYSTG 675  
Qy 541 QVSVEIEWELQKNSKRWNPEVQYTSNYAKSANVDSVDNNGLYTSPRIGRYLTRPL 599  
Db 676 QVSVEIEWELQKNSKRWNPEIQYTSNYKNSVNVDFTVDTNGVYSEPRPIGRYLTRNL 734

RESULT 15

AAG65794  
ID AAG65794 standard; Protein; 533 AA.

XX AAG65794;  
XX 11-FEB-2002 (first entry)  
XX Adeno-associated virus 2 (AAV-2) major coat protein VP3.

XX Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
KW inverted terminal repeat; nontropic; neuroprotective; antianemic; ITR;  
KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
XX major coat protein; AAV-2; VP3.

XX Adeno-associated virus 2.

OS WO200168888-A2.

PN XX

XX 20-SEP-2001.

XX 13-MAR-2001; 2001WO-US07927.

XX 14-MAR-2000; 2000US-189110P.

XX (NEUR-) NEUROLOGIX INC.

XX Xiao W, During MJ;

XX WPI; 2001-596912/67.

DR N-PSDB; AAI66974.

XX Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a  
PT chimeric capsid having one non-native amino acid sequence and a desired  
PT transgene -

XX Disclosure; Page 51; 53pp; English.

XX The invention provides a recombinant viral vector (RVV) comprising a  
CC chimeric capsid (I) having at least one non-native amino acid sequence,  
CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
CC or their combination, and a transgene flanked 5' and 3' by inverted  
CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
CC combination. The RVV is useful for improving gene therapy in a subject  
CC with a disorder, and for increasing the efficiency of entry into a cell,  
CC which involves producing (I) encapsulating a viral vector, and contacting  
CC a cell with RVV having (I) such that (I) binds to an attachment site on  
CC the cell surface and permits the vector to enter the cell efficiently.  
CC A pharmaceutical composition comprising RVV with (I) containing a  
CC transgene sequence associated with a disease or a disorder such that  
CC expression of the transgene would result in amelioration of the disease  
CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polyneuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences  
CC are used in the construction of a chimeric vector.

XX Sequence 533 AA;

XX SQ

Query Match 76.3%; Score 2481.5; DB 22; Length 533;  
Best Local Similarity 83.3%; Pred. No. 2.8e-192;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
Qy 66 MASGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTTTRTWALPTNNHLYKQISS 125  
Db 1 MATGSGAPMADNNEGADGVGNSGNNHCHDSTWMDRVITTTTRTWALPTNNHLYKQISS 60  
Qy 126 ASTGASNDNHFGYSTPMGYDFNRFCHFSRDRQRLNNNNWGRPRKLNKFLNIOVK 185  
Db 61 QS-GASNDNHFGYSTPMGYDFNRFCHFSRDRQRLNNNNWGRPRKLNKFLNIOVK 119  
Qy 186 EVTTNDGVTTIANLNTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVPMIPIQYGLTILN 245  
Db 120 EVTQNDGTTIANLNTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVPMVPIQYGLTILN 179  
Qy 246 NGSQAVGRSSPYCLEYFPFSQMLRTGNFTFSYTFPEVPHSSYAHQSQSLDRMLNPLIDQY 305  
Db 180 NGSQAVGRSSPYCLEYFPFSQMLRTGNFTFSYTFPEVPHSSYAHQSQSLDRMLNPLIDQY 239  
Qy 306 LYLLNRTQNSGSAQNKDLLFSRGSFAGMSVQPKWLPGPCYRQORVSKTKTDNNNSNFT 365  
Db 240 LYLSRTNTPSGTTTQSRLOFSQAGASDIRDSRWLPGPCYRQORVSKTSADNNNSYS 299  
Qy 366 MTGASKYNLNGRESINFEKTAMASHKDEDEKFFPPMSGVMI FGKESAGASNTALDNVMTD 425  
Db 300 MTGATKYHLNGRDSLNVNPGPAMASHKDEEKFPPQSGVLI FGKQSGSEKTNVDIEKVMITD 359  
Qy 426 EEEIKATNPVATERGTVAVNPFSSSTDPATGCDVHAMGALPGMWQDRDVIYLOGPIWAKI 485  
Db 360 EEEIRTNVPATEQYGSVSTNLQRNQAAATADVNTQGVLPGMVWQDRDVIYLOGPIWAKI 419  
Qy 486 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFSFITYSTQGVSV 545  
Db 420 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPSTTFSAAKFASFIQYSTQGVSV 479  
Qy 546 IEWELQKNSKRWNPEVQYTSNYAKSANVDFTVDTNGVYSEPRPIGRYLTRPL 599  
Db 480 IEWELQKNSKRWNPEIQYTSNYKNSVNVDFTVDTNGVYSEPRPIGRYLTRNL 533

Search completed: January 21, 2004, 16:04:19  
Job time : 37.5361 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:00:42 ; Search time 14.1017 Seconds  
(without alignments)  
4084.983 Million cell updates/sec

Title: US-09-807-802A-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NNGLYTEPRPIGTRYLTRPL 599  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

				SUMMARIES	
Result No.	Score	Query Match	Length DB ID	Description	
1	2177.5	67.0	504 1 VCPV3A	coat protein - ade	
2	1701	52.3	732 2 S52210	coat protein vp1	
3	556	17.1	673 1 VCPVB5	coat protein vp1	
4	495	15.2	781 1 VCPV19	coat protein vp1	
5	258.5	8.0	723 1 VCPVPP	coat protein vp1	
6	254.5	7.8	729 1 VCPVNA	coat protein vp1	
7	254.5	7.8	729 1 A60006	coat protein vp1	
8	244	7.5	722 1 VCPVV2	coat protein vp1	
9	234	7.2	587 1 B44276	coat protein vp1	
10	216	6.6	716 1 VCPV2M	coat protein vp1	
11	216	6.6	727 1 VCPV1F	coat protein vp1	
12	214.5	6.6	718 1 VCPVIM	coat protein vp1	
13	214	6.6	722 1 VCPVME	coat protein vp1	
14	209	6.4	727 1 VCPVFP	coat protein vp1	
15	208	6.4	737 1 VCPVCP	coat protein vp1	
16	204	6.3	748 1 VCPVCP	coat protein vp1	
17	198	6.1	584 2 S49594	capsid protein vp2	
18	192	5.9	722 1 VCPVCN	coat protein vp1	
19	191	5.9	702 1 VCPVAP	coat protein vp1	
20	143.5	4.4	648 2 S50856	whn protein - rat	
21	129	4.0	1072 2 A86827	hypothetical prote	
22	123.5	3.8	690 2 AB0124	probable TonB-depe	
23	123	3.8	1142 2 T37455	enamelin precursor	
24	122.5	3.8	880 1 SYBSVS	valine-tRNA ligase	
25	121	3.7	635 2 F96660	nuclear pore compl	
26	121	3.7	1113 2 S28925	related to glucan	
27	120.5	3.7	931 2 T49710	transcription fact	
28	119	3.7	667 2 A41311	hypothetical prote	
29	118.5	3.6	1180 2 E86719		

30 118 3.6 739 2 TS2289 probable transketo  
31 116.5 3.6 642 1 S34416 transcription fact  
32 116.5 3.6 1338 2 T30565 MAP kinase kinase  
33 115.5 3.6 345 1 G97024 probable phosphoes  
34 114 3.5 956 2 T08144 myrosinase-binding  
35 114 3.5 2271 2 F90073 hypothetical prote  
36 113.5 3.5 2817 2 B97033 uncharacterized pr  
37 113 3.5 1071 2 E85343 hypothetical prote  
38 112.5 3.5 666 2 I52648 class A helix-loop  
39 112.5 3.5 3078 2 T28432 variant-specific s  
40 111 3.4 1777 2 T34369 hypothetical prote  
41 110.5 3.4 655 1 ALKBG celluomaltodextrin  
42 110.5 3.4 1742 2 T17120 cellulase (EC 3.2.  
43 110 3.4 868 2 A82515 conserved hypothet  
44 110 3.4 1296 2 C82521 hemolysin-type cal  
45 110 3.4 2825 2 T14271 Doc4 protein, stre

ALIGNMENTS

RESULT 1

VCPV3A

coat protein - adeno-associated virus type 2

C:Species: adeno-associated virus type 2

C:Date: 05-Apr-1983 #sequence\_revision 05-Apr-1983 #text\_change 16-Jul-1999

C:Accession: A03698

R:Srivastava, A.; Lusby, E.W.; Berns, K.I.

J: Virol. 45, 555-564, 1983

A:Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.

A:Reference number: A03694; MOID:83164299; PMID:6300419

A:Accession: A03698

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-504 <SRI>

A:Cross-references: EMBL:J01901; NID:G209616; PID:AAA42376.1; PID:G209621; EMBL:M12405.

C:Superfamily: adeno-associated virus coat protein

C:Keywords: coat protein

Query Match 67.0%; Score 2177.5; DB 1; Length 504;  
Best Local Similarity 82.5%; Pred. No. 8.9e-142;  
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;

Qy	66	MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYNHLYKQISS	125
Db	1	MATGSGAPMADNNEGADGVGNAGNWHCDSTWMDRVITTTSTRTWALPTYNHLYKQISS	60
Qy	126	ASTGASNDNHYGYSTPWGYFDNRFCHFSRDNQORLNNWGFPRKLNLFNIOVK	185
Db	61	QS-GASNDNHYGYSTPWGYFDNRFCHFSRDNQORLNNWGFPRKLNLFNIOVK	119
Qy	186	EVTINDGVTTIANNTSTVQVFSSEYOLPYVLGSAHQCLPPFPADVFMIPOYGYLTIN	245
Db	120	EVTQNDGTTIANNTSTVQVFTDSEYOLPYVLGSAHQCLPPFPADVFMIPOYGYLTIN	179
Qy	246	NGSQAVGRSSFCYCLEYFFSQMLRTGNFTSYTPEVPFHSSYAHQSQSLDRMLNPLIDQY	305
Db	180	NGSQAVGRSSFCYCLEYFFSQMLRTGNFTSYTPEVPFHSSYAHQSQSLDRMLNPLIDQY	239
Qy	306	LYLARTQNSGSAQNKDLLFSRGSFAGMSVQPKWLPGPCYRQORVSKTKTDNNNSNFT	365
Db	240	LYLSTRTPSGTTTQSLRQFSQAGASDIRDQSRWMLPGPCYRQORVSKTSADNNNSYS	299
Qy	366	WTGASKYNLNGRESINPGTAMASHKDDKFFPMGSMVMI FCKEAGASANTALDNVMTD	425
Db	300	WTGATKYHLNGRDSLVP--AVASHKDDBEKFFPGSGVLI FCKQSEKTNVNIKVMITD	357
Qy	426	EBEIKATNPVATERFGTVAVNPFQSSSTDPATCDVHAMGALPGMWQDRDVLQGPWAKI	485
Db	358	EBEIGTNPVATEQYGVSVSTNLQRNQAATADVNTQGLPGMWQDRDVLQGPWAKI	417
Qy	486	PHTDGHFHPSPLMGGFGLKNPPPTQLIKNTVPANPPAEFSATKFSPTQYSTG	540

Db 418 PHTDGHFSPLMGGFGLKHPPIIKNTVPANPSTTFSAKPFITQYSTG 472

RESULT 2

S52210

coat protein VP1 - muscovy duck parvovirus

N:Alternate names: VP1 protein

C:Species: muscovy duck parvovirus

C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-2000

C:Accession: S52210

R:Zadori, Z.; Erdel, J.; Nagy, J.; Kisari, J.

submitted to the EMBL Data Library, September 1993

A:Reference number: S52209

A:Accession: S52210

A:Molecule type: DNA

A:Residues: 1-732 <ZAD>

A:Cross-references: EMBL:X75093; NID:9609091; PIDN:CAAS2984.1; PID:9609093

A:Experimental source: strain FM

C:Genetics:

A:Gene: VP1

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

Query Match 52.3%; Score 1701; DB 2; Length 732;  
Best Local Similarity 53.0%; Pred. No. 7.8e-109;  
Matches 325; Conservative 83; Mismatches 165; Indels 40; Gaps 11;

QY 6 KPVQSPQSPQSSGIGK-TQQPAKKELNFGQTGDSSEVPDPQLG-----E 53  
DB 141 EEPVNTAPAKSS-----GKLTDDHPVTKPKLSE-----ENSPSPNSGGEASAAATEGSE 192

QY 54 PPATPAAVGPTTMAAGGAPMADNNEGADVGNASGNHCHDSTWLGDRVITTTSTRTWALP 113  
DB 193 PVAAP-----NMAEGSGAMGDSAGGADVGNASGNHCHDQWLGDTVIKTTITWLP 246

QY 114 TNNHLYKOISASTGASNDNHYFGYSTPWGYFDNRFHCHFSPRDWORLNNNGFRPK 173  
DB 247 SYNHHYQAITSQTNPDSN-TQYAGYSTPWGYFDNRFHCHFSPRDWORLNNHGWIRPK 305

QY 174 RLNFKLQNKVQEVTTNDGVTIANN-LTSTVQVSDSEYQLPYVLGSAHQCLPPFPADV 233  
DB 306 ALKFLIFNVQEVTTQDOTKIANNLTSTIQIFDNEHQLPYVLGSAHQCLPPFPADV 365

QY 234 FMIPQGYLTN---NGSAQVRSFYCYLPPSQMLRTGNFTSYTPEEVPFHSVAH 290  
DB 366 YALPQYGYCTMTNQSAGRNDSAFYCYLPPSQMLRTGNFTSYTPEEVPFHSMEAH 425

QY 291 SQSLDLNPLIDQYLYLNRNTQNSGSAONKDLFLSRGSPAGMSVQPKWLPQPCYRQQ 350  
DB 426 SQDLRLNPLIDQYLYLNFSEV-NGGRNAQ-----FKKAVKGAFAAGMGRNWLPGPKLLDQ 479

QY 351 RVSKTK--TDNNNSNFTWGTASKYLNNGRESINPGTAMASHKDDKDFPMSGVMIQK 408  
DB 480 RVRAVSGGTDNANWSINSKGNKVLKDRYLLQPGPVATTHTEQASVPQAIIIGAK 539

QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHWAGALP 466  
DB 540 DPNRSGSTLAGISDIWVTDEQEIAPTNGVGRPYGLTVTNEQTITAPTNAEVLVGLP 599

QY 467 GNVMDRDVYVLOGPIWAKIPHTDGHFSPLMGGFGLKHPPIIKNTVPANPAPFS 526  
DB 600 GNVMDRDVYVLOGPIWAKIPHTDGHFSPLMGGFGLKHPPIIKNTVPANPAPFS 659

QY 527 ATKFASFITQYSTGQVGSVIEWELQENSKRNWPEVQYTSNYAKSANVDFTVDNNGLYTE 586  
DB 660 NQWNSYITQYSTGQVGSVIEWELQENSKRNWPEVQYTSNYAKSANVDFTVDNNGLYTE 719

QY 587 PRPIGTRYLTRPL 599  
DB 720 DRLIGTRYLTRPL 732

RESULT 3

VCVPB5

coat protein VP1 - bovine parvovirus

N:Contains: coat protein VP2

C:Species: bovine parvovirus

C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999

C:Accession: A26104

R:Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.

J. Virol. 60, 1085-1097, 1986

A:Title: Complete nucleotide sequence and genome organization of bovine parvovirus.

A:Reference number: A26104; MUID:87061184; PMID:3783814

A:Accession: A26104

A:Molecule type: DNA

A:Residues: 1-673 <CHE>

A:Cross-references: EMBL:M14363; NID:G333454; PIDN:AAB59847.1; PID:9808805

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

F:138-673/Product: coat protein VP2 #status predicted <VP2>

Query Match 17.1%; Score 556; DB 1; Length 673;  
Best Local Similarity 25.0%; Pred. No. 2.7e-30;  
Matches 160; Conservative 98; Mismatches 239; Indels 144; Gaps 19;

QY 1 TAPKKRPPVQSPQSPQSSGIGK-TQQPAKKELNFGQTGDSSEVPDPQLGEP--PATP 58  
DB 100 TSKGGDRALKRKLFAFSNKGAKKANREPAPSTSNQONMEVSNIDIPNDEAGNQPIELATR 159

QY 59 AAVGPTTMAAGGAPMADNNEGADVGNASGNHCHDSTWLGDRVITTTSTRTWALPYYNNH 118  
DB 160 SVUGSGVGGG-----RGSGVGYSTGGTGTIPSENIIVVTXTRQFICDIKNGH 211

QY 119 LYKOISASTGASNDNHYFGYSTPWGYFDNRFHCHFSPRDWORLNNNGFRPKLNFK 178  
DB 212 LYKS-EVLNTGDTAHRQY-AITTPWSYFNQYSSHSFSPNDQWHLVNDYERFRPKAMIVR 269

QY 179 LFNQVKEVTTNDGVTIANN-LTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVMP 237  
DB 270 VYNLQIKQIMTDGAMGTVYNNDLTAGMHIIFCDGGRYPYVQHPWDDQCMPELPSIWELP 329

QY 238 QYGLTL-----NNGSAQVRS-----SFVLEYFPFSQMLRTGNFTSYTPEEVPFHS 286  
DB 330 QYAIPIAPISVVDNNTTVEEHLKGVPLYMLENSDHEVLNNG----- 373

QY 287 SYAHQSRLDNPLIDQYLYLNRNTQ-----QSGSAQNKDLLFSRQSP----- 331  
DB 374 -----RIYRIYIQLWRLMDRKQHHQHSHASDDVQSTGQKQKLLIQTQPNKORF 424

QY 332 AGMSVQPKWLPQPCYRQVRVSKTKTDNNNSNFTWGTASKYLNNGRESINPGTAMASHK 391  
DB 425 QNAALRTSNMMSGP-----GIARTGHTNATLOTQSGALVTWVT 462

QY 392 DDEKDFPMSGV-----IFGKE-----SAGASNTALDNVMTDE 426  
DB 463 NGAD-----VSGVRAVRVGYSTDPYGGQPSDLRLRYASASAAEQQNPLEN----- 512

QY 427 BEIKATNPVATERFGTVAVNFQSSSTDPATGDVHWAGALPQWVQDRDYLQGTWAKIP 486  
DB 513 -----AARHTTREATKLTITGNSGADGDYKEMWMLPQNMWDSAPISRYNPVWVKVP 564

QY 487 HTDGHFSPLMGGFGLKHPPIIKNTVPANPAPFSATKFAFITQYSTGQVSV 544  
DB 565 RVNRTLLDQDGGIPSHPPGTIFIKLARIPVPCNGD-----SFLNIYVTGQVSC 615

QY 545 IEWELQENSKRNWPEVQYTSNYAKSANVD-FTVDNNGLY 584  
DB 616 EVWVEVEKRGTKNRPEYMHSS---ATNMSVDAYTINNAGVY 653

RESULT 4

VCVP19

coat protein VP1 - parvovirus B19 (strain Au)

C:Species: parvovirus B19

A:Note: host Homo sapiens (man)

C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999

[illegible]

```

RESULT 6
VCPVNA
coat protein VP1 - porcine parvovirus (strain NADL-2)
N:Contains: coat protein vp2
C:Species: porcine parvovirus
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: B33743; D48472
R:Vasudevacharya, J.; Basak, S.; Srinivas, R.V.; Compans, R.W.
Virology 173, 368-377, 1989
A:Title: Nucleotide sequence analysis of the capsid genes and the right-hand ter
A:Reference number: A33743; MUID:90085785; PMID:2596019
A:Accession: B33743
A:Molecule type: DNA
A:Residues: 1-729 <VAD>
A:Cross-references: GB:M32787; NID:G3322983; PIDN:AAA46917.1; PID:G3322985

```

R; Bergeron, J.; Menezes, J.; Tijssen, P.

Virology 197. 86-98, 1993

A; Title: Genomic organization and mapping of transcription and translation products of porcine parvovirus

A; Reference number: A48472; MUID: 94025614; PMID: 8212598

A; Accession: D48472

A; Molecule type: DNA

A; Residues: 11-729 <BER>

A; Experimental source: NADL-2, ATCC VR-742

A; Note: sequence extracted from NCBI backbone (NCBI:138789, NCBIP:138794)

C; Genes: 10/1

C; Superfamily: parvovirus coat protein

C; Keywords: coat protein; glycoprotein

F; 151-729/Product: coat protein VP2 #status predicted <VP2>

F; 172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #status predicted <VP2>

Query Match 7.8%; Score 254.5; DB 1; Length 729;  
Best Local Similarity 22.7%; Pred. No. 1.5e-09;  
Matches 152; Conservative 94; Mismatches 272; Indels 151; Gaps 34;

QY 9 VEQSPQEPDSSSGIGK-----TGQOPAKKRLNFGQTGDSVDPDPQLGEPPATP 58

DB 109 VRSPRKHPGSKPPGRAPRHIFINLAKKAKGTSTNSNSMSENVEQHNPINAGTEL- 167

QY 59 AAVGPTTMSAGGAPMADNNEGADGVGNAG--NMHCDSTWLG--RVITSTRTWALP 113

DB 168 SATGNSGGGGGGG-----GGRGAGGVSTGTENNQTEFOYLGEGVLRIITAHASRLIHLN 223

QY 114 TYNHLYKQIS--SASTGAS-----NDNHFGYSTPWGYDFNRFCHFSRPRDQRLINN 167

DB 224 MPHEHYTKRIHVLNSESAGVQVQDDAHTQMTWPSLIDANAWGVNFPADWQLISNNM 283

QY 168 WGFPRKRLNFKLNIOQKEVT---TNDGVTTIANNTSTVQVPSDSEYQLPYVLGSAHQ 224

DB 284 TEINLVSFEQEIFNVVLKTTESATSPPTKIYNNDLTASLMVALDTNNTLPTPAAPRSE 343

QY 225 CLPPFPADVPMIPQGYL-----TLNNGSAVGRS-----SFYCLE-YFPS 264

DB 344 TLGFYPLPTKPTQYRYLSLRNLPPTVTGOSQOITDSIQGLHSDIMFYTIENAVPI 403

QY 265 QMLRTGNPF-TFSYTFEEVFPFSSVSAHSQSLDR-----LMNPLI--DOYLYLNRTO 313

DB 404 HLLRTGDEFSTGIYHFDTKPL--KLTHSWQTNRSGLPPLKLTETPTTEGQDHPGTLPAAN 461

QY 314 NQSGAQNKDLFSRSGPAGMSVQPKWLPQPCYQQRVSKTKTDNNNSNFTWTGASKYN 373

DB 462 TRKGYHQITNNSYTEAT---AIRP-----AQVYNTPYNFEYSNGGPP- 502

QY 374 LNGRESIINP--GTAMASHKODEDKFPFMSGVMIFFKESAGASANTALD---NVNITDEE 428

DB 503 -----LTPIVPTADTQYNDDEPN-----GAIRFTMDYQHGLHTTSQE 540

QY 429 IK--ATNP-----VATERFGTVA-VNFQSS-----TDPATG--DVHAMGALP----- 466

DB 541 LERYTFNQSKCGRAPKQFNQQAFLNLENTNNTGLLPSPDIGKPNMHPMNTLNTYGPL 600

QY 467 GMVQDRDVLQGPWAKIPIHTD--GHFHPSPLMGGFGLK--NPPQILIKNTVPANPFA 523

DB 601 TALNNTAPVFPNGQIDKELDTLKLRLH---VTAPVCKNPPGQLFVKIAP---NLTD 654

QY 524 EPSA-TKFASTFYQTSQVSVIEWELQKENSCKRNPEVQYTSNYSKANSVDFVNDNG 582

DB 655 DFNADSPQPRITYSNFWKGLTFTAKVRSSNMNPIQOHTT-----TAENIG 704

QY 583 LYTEPRPIG 591

DB 705 NYI-PTNIG 712

RESULT 7

A60006

coat protein VP1 - porcine parvovirus (strain 90HS)

N; Contains: coat protein VP2

C; Species: porcine parvovirus

C; Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 30-Sep-1993

C; Accession: A60006

R; Sakurai, M.; Nishimori, T.; Ushimi, C.; Nakajima, H.

Virus Res. 13, 79-86, 1989

A; Title: Nucleotide sequence of capsid protein gene of porcine parvovirus.

A; Reference number: A60006; MUID: 89319168; PMID: 2750278

A; Accession: A60006

A; Molecule type: DNA

A; Residues: 1-729 <SAK>

C; Superfamily: parvovirus coat protein

C; Keywords: coat protein; glycoprotein

F; 151-729/Product: coat protein VP2 #status predicted <VP2>

F; 172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #status predicted <VP2>

Query Match 7.8%; Score 254.5; DB 1; Length 729;  
Best Local Similarity 22.7%; Pred. No. 1.5e-09;  
Matches 151; Conservative 95; Mismatches 277; Indels 141; Gaps 33;

QY 9 VEQSPQEPDSSSGIGK-----TGQOPAKKRLNFGQTGDSVDPDPQLGEPPATP 58

DB 109 VRSPRKHPGSKPPGRAPRHIFINLAKKAKGTSTNSNSMSENVEQHNPINAGTEL- 167

QY 59 AAVGPTTMSAGGAPMADNNEGADGVGNAG--NMHCDSTWLG--RVITSTRTWALP 113

DB 168 SATGNSGGGGGGG-----GGRGAGGVSTGTENNQTEFOYLGEGVLRIITAHASRLIHLN 223

QY 114 TYNHLYKQIS--SASTGAS-----NDNHFGYSTPWGYDFNRFCHFSRPRDQRLINN 167

DB 224 MPHEHYTKRIHVLNSESAGVQVQDDAHTQMTWPSLIDANAWGVNFPADWQLISNNM 283

QY 168 WGFPRKRLNFKLNIOQKEVT---TNDGVTTIANNTSTVQVPSDSEYQLPYVLGSAHQ 224

DB 284 TEINLVSFEQEIFNVVLKTTESATSPPTKIYNNDLTASLMVALDTNNTLPTPAAPRSE 343

QY 225 CLPPFPADVPMIPQGYL-----TLNNGSAVGRS-----SFYCLE-YFPS 264

DB 344 TLGFYPLPTKPTQYRYLSLRNLPPTVTGOSQOITDSIQGLHSDIMFYTIENAVPI 403

QY 265 QMLRTGNPF-TFSYTFEEVFPFSSVSAHSQSLDR-----LMNPLI--DOYLYLNRTO 313

DB 404 HLLRTGDEFSTGIYHFDTKPL--KLTHSWQTNRSGLPPLKLTETPTTEGQDHPGTLPAAN 461

QY 314 NQSGAQNKDLFSRSGPAGMSVQPKWLPQPCYQQRVSKTKTDNNNSNFTWTGASKYN 373

DB 462 TRKGYHQITNNSYTEAT---AIRP-----AQVYNTPYNFEYSNGGPP- 502

QY 374 LNGRESIINP--GTAMASHKODEDKFPFMSGVMIFFKESAGASANTALD---NVNITDEE 431

DB 503 -----LTPIVPTADTQYNDDEPN-----GAIRFTMDYQHGLHTTSQE 545

QY 432 TNP-----VATERFGTVA-VNFQSS-----TDPATG--DVHAMGALP-----GMVQ 471

DB 546 FNPQSKCGRAPKQFNQQAFLNLENTNNTGLLPSPDIGKPNMHPMNTLNTYGPLTALNN 605

QY 472 DRDVLQGPWAKIPIHTD--GHFHPSPLMGGFGLK--NPPQILIKNTVPANPFA 527

DB 606 TAPVFPNGQIDKELDTLKLRLH---VTAPVCKNPPGQLFVKIAP---NLTD 659

QY 528 TKFASTFYQTSQVSVIEWELQKENSCKRNPEVQYTSNYSKANSVDFVNDNG 587

DB 660 SPOQPRITYSNFWKGLTFTAKVRSSNMNPIQOHTT-----TAENIG 708

QY 588 RPIG 591

DB 709 TNIG 712

RESULT 8

VCPW2

coat protein VP1 - parvovirus H1

C; Species: parvovirus H1

A; Note: host Homo sapiens (man)







A:Accession: B29510  
 A:Molecule type: DNA  
 A:Residues: 1-143, 'A', 145-718 <AST>  
 A:Cross-references: EMBL:M12032  
 C:Superfamily: parvovirus coat protein  
 C:Keywords: coat protein

Query Match 6.6%; Score 214.5; DB 1; Length 718;  
 Best Local Similarity 21.0%; Pred. No. 8.1e-07;  
 Matches 142; Conservative 273; Mismatches 273; Indels 171; Gaps 30;

QY 8 PVESQSPQSSSGIGKTG-----QPAKRLNFG---QTGDSSEVPDPQLG 52  
 DB 83 PKLATSEP-GTSGVSRAGKTRPPAYIFINQARAKKLTSSAAQSSQTMDSGTSPDQ 141  
 QY 53 -----EPPATPAAGVPTTMASSGGAPMADNNEGADGVGNAGNWHDS--TWLGP--R 101  
 DB 142 GNGVHSNARVERAADPG--GSGG-----GSGGGVGVSSTGYNQTHYRFLGQWVE 193  
 QY 102 VITSTRTWALPTNNHLYKQI-----SSASTGASNDNHFYCYSTPMGYFDNRFHCH 154  
 DB 194 ITALATRLVHLNPKSENYCRIVHNTTDTSVKGNMAKDDAHEQIWTWPSLVDANAGVW 253  
 QY 155 FSPRDQRLNNWGRPKRLNFKLNIQKVEVTND-----GVTTANNLTSTVQVFSOS 210  
 DB 254 LQPSDMQYICNTMSQLNLVSLDQIEFNVLKTVTEODSGGCAIKIYNNDLTACMVAVDS 313  
 QY 211 EYOLPYVLGSAHQCLPPFPADVFMIPQGY-----LTLNN-----GSAV 251  
 DB 314 NNLTPTPAANSMETLGFYFWKPTIASPYRYFCVDRDLSTVYENQEGTIEHNVMTGPKG 373  
 QY 252 GRSSFCYLEYFP--QMLRTGNF--TFSYTFEEVPHSSVAHSQSLSRLMN--PLIDQYLY 307  
 DB 374 MNSQFFTIENQOITLLRTGDEPATGTYYPDTNPV--KLTHWTQNRQLGQPLLSTF-- 429  
 QY 308 YLNRTQNSQSAQKLLFSPGSPAGMSVQPKWL-----LTAQSRHGATQMVNVSEAIRTPAQVGFQCPHNDFEASRA 481  
 DB 430 --PEADTDAGT-----LTAQSRHGATQMVNVSEAIRTPAQVGFQCPHNDFEASRA 481  
 QY 344 GPCYRQORVSKTKTDNNNSNFTWGASKYNLNGRESLIINPGTAMASHKDDKDF----- 397  
 DB 482 GP-PAAPKVPADVQTGVDREANGSVYSYKQGEWAHAGPAPERYTWDETNGSGRDT 540  
 QY 398 ---FPMGVMIFOKESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQS---SST 452  
 DB 541 RDGFIOASAPLVVPPPLNGILTNA-----NPIGTKN-----DIHFSNVFNSV 581  
 QY 453 DPATGDVHANGALPGMWQDRDVLQGPWAK---IPHTDGHFHPSPLMGGFGLKNPPQ 509  
 DB 582 GLPTAFSH-----PSPVYPOQIWDKELDLEHKPELHITAPFV---CKNNAPOG 627  
 QY 510 ILIKNTNP--VPANPPAEFSATKFSFTQYSTQGSVEIEMELQKENSKRWNPEVQYTSN 567  
 DB 628 MLVRLGPNLTQDYPNGATLSRIVTYGTFFPKWGLTHRAKIRA-----NTTNVPIYO---- 679  
 QY 568 YAKSANVDFTVONNG 582  
 DB 680 -----VSVEDNG 686

## RESULT 13

## VCPVME

coat protein VP1 - mink enteritis virus (strain Abashiri)  
 N:Contains: coat protein VP2  
 C:Species: mink enteritis virus, MEV  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 12-Apr-1996  
 C:Accession: B38350  
 R:Kariatsumari, T.; Horiuchi, M.; Hama, E.; Yaguchi, K.; Ishiguro, N.; Goto, H.; Shinagawa, J. Gen. Virol. 72, 867-875, 1991  
 A:Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the mink enteritis virus  
 A:Reference number: A38350; MUID:91202123; PMID:2016597  
 A:Accession: B38350  
 A:Molecule type: DNA

A:Residues: 1-722 <KAR>

A:Cross-references: GB:D00765

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

F:139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.6%; Score 214; DB 1; Length 722;  
 Best Local Similarity 20.2%; Pred. No. 8.8e-07;  
 Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;

QY 7 RPVESQSPQSSSGIGKTGQPAKRLNFGQTGDSSEVP-----DPQIAGEPPA--TPAA 60  
 DB 105 KTKRSKPPHIFINLAK-----KKAGAGQKRDNLAPMSDGAQVDPGGQPAVRNERA 158  
 QY 61 VCPPTMASGGAPMADNNEGADGVGNAGNWH-----CDSTWLGDRVITTTSTRWALP 113  
 DB 159 TSGNGSGGG-----GGSGGVGISTGTFFNQTEFKFLENGWV--EITANSSRLVHLN 210  
 QY 114 TYNNHLYKQI-----SSASTGASNDNHFYCYSTPMGYFDNRFHCHFSPRDWORLINN 166  
 DB 211 MPESENYKRVVNNMDKTAVKGNMALDDTHQIVTPWSLVDANAGVWPNPGDQLIVNT 270  
 QY 167 NMGFPKRLNFKLNIQKVEVT---TNDGVTIANNLTSTVQVFSQSEYQLPYVLGSAHQ 223  
 DB 271 MSELHLVSFEQIEFNVLKTVSESATQPPTKVYNNDLTASLMVALDSNNTMBTTPAAMRS 330  
 QY 224 GCL-----PPPPA-----DVEWIPQYGLTLNNGSQAVGRS-----SFYCL 259  
 DB 331 ETLGYFWKPTIPTWRYFQWDRTLIPSH-----TGTSGPTTNIYHGTDDPDDVQFYTI 384  
 QY 260 E-YFESQMLRTGNF--TFSYTFEEVPHSSVAHSQSLSRLMN--PLIDQYLYLNRTOQ 315  
 DB 385 ENSVPVHLLRTGDEPATGTFPDKCP--CRLTHWTQNRALGLPP-----PLNSLPQS 435  
 QY 316 SGSAQKOLLFSPGSPAGMS-----VOPKN-WLPGCYRQORVSK-----354  
 DB 436 EGATNFGDIGVQDQKRRGVTMGNTDYITEATIMPAEYGVGAPYYSFEASTQGGPKTI 495  
 QY 355 -----TKTDNNNSNFTWGASKY---NLNGRESLIINPGT-----AMASHKDDKFFPMS 401  
 DB 496 AAGRGAQTDENQAA---DGDPRYAFGRHQGQKTTTTTETPERPTYIAHQDT----- 544  
 QY 402 GVMIFKESAGASNTALD--NVMITDEEIKATNPVATERFGTVAVNFQSSDTPATGDVH 460  
 DB 545 -----GRYPAGDWIQINFNPLPVTNDVLLPTDPIG---GKTGINY--TNIFNTYGPLT 593  
 QY 461 ANGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPLMGGFGLK-----NPP 507  
 DB 594 ALNNVP-----PVYPNGQIWDKEFDTD-----LKPRLHYNAPFVQCNQNC 633  
 QY 508 PQLIKNTNPVAN---PPAEFSATKFSFTQYSTQGSVEIEMELQKENSKRWNPEVQY 564  
 DB 634 GOLFKVAPNLNEDVDPASANMSR---IVTYSDFWKGKLVFKAKLRASHTWNPQIM 689  
 QY 565 TSNYAKSANVDFTVDNNGLYTEPRPIG 591  
 DB 690 SIN-----VDNQFNYL-PNNIG 705

## RESULT 14

## VCPVFP

coat protein VP1 - feline panleukopenia virus (strain 193)  
 N:Contains: coat protein VP2  
 C:Species: feline panleukopenia virus, FPLV  
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 12-Apr-1996  
 C:Accession: B36608  
 R:Martyn, J.C.; Davidson, B.E.; Studdert, M.J.  
 J. Gen. Virol. 71, 2747-2753, 1990  
 A:Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvovirus  
 A:Reference number: A36608; MUID:91073139; PMID:2174965  
 A:Accession: B36608  
 A:Molecule type: DNA  
 A:Residues: 1-727 <MAR>

A:Cross-references: GB:X55115  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F:144-727/Product: coat protein VP2 #status predicted <VP2>

```
Query Match      6.4%; Score 209; DB 1; Length 727;
Best Local Similarity 20.1%; Pred. No. 2e-06;
Matches 138; Conservative 101; Mismatches 260; Indels 188; Gaps 34;

QY  7 RPVEQSPQEPSSGIGKGTQQAQKRLNFGTGDSESV-----DPQLGEPPEA--TPAA 60
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 KPTKRSKPPPHIFINLAK-----KKKAGAGQVKRDNLAPMSDGAQVDPGGQPAVRNERA 163
QY  61 VQPTTMASGGAPMADNNEGADGVGNASGNWH-----CSTWLGDRVITSTRTWALP 113
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 TSGNGSGGG-----GGSGGVIGISTGFNNQTEPKFLENGWV--EITANSRLVHLN 215
QY  114 TYNHLYKQI-----SSASTGASNDNHFGYSTPMGYDFPNRHFCHFSPRDQRLINN 166
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 MPESYKRVVNNMDKTA VGNMALDDIHQIVTPWLSLVDANAWGVNFPNGDQWLIVNT 275
QY  167 NWGPRPKRLNPKLNIQVKEVT---TNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHQ 223
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 MSELHLVSFEQEIFNVVLKTVSESATQPTKYNNDLTASLWALDSSNTMPFTPAAMRS 335
QY  224 GCL-----PPPPA-----DVFMIPQGYLTLLNGSQAVGRS-----SFYCL 259
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 ETLGFPYKWTPTTPWRYFQWDRTLIPSH-----TGTSGTPTNVYHGTDPDDVQFYTI 389
QY  260 E-YFPSQMLRTGNF-TFSYTFEEVPHSSYAHQSQSLRLMN--PLIDQYLYLNRNQ 315
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 390 ENSVPVHLLRTGDEFATGTFDFCKP--CRLTHTWQTNRALGLP-----FLNSLPQS 440
QY  316 SGAQNKOLLPSRGSFAGMS-----VQPKN-WLPGCYRQORVSK----- 354
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 EGATNFGDIGVQDQKRGVQMGNTDYITEATMPAEVGSAPYSPSEASTQGFKTP 500
QY  355 -----TKTDNNNSFTWTGASKY---NLNGRESIINPGT---AMASHKDEDEKFFPMS 401
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 AAGRGGAQTDENQAA---DGDPRYAFGRHQGKTKTTTGTPTERTFYIAHQDT----- 549
QY  402 GVMIFKESAGASNTALD-NVMTDEEIKATNPVATERFGTVAVNFSSSTDPATGDVH 460
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 -----GRYPEGDWIQINFNLPVNDNVLLPTDPIG---GKTGINY--TNIFNYGFLT 598
QY  461 AMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGFGFLK-----NPP 507
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 599 ALNNVP-----PVYPNGQIWDKEFTD-----LKPRLLHVNAPFVQNNCP 638
QY  508 PQILIKNTFVPAN---PPAEFSATKPFASFIQYSTQSVSEIWEIWEIWEIWEIWEIWEI 564
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 639 GQLFVKVAPNLNTEYDPPDASANMSR---IVTYSDFWPKWKGKLVFKAKLRASHTWNP 694
QY  565 TSNYAKSANVDFTVNNGLYTEPRPIG 591
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 695 SIN-----VDNQFNIV--PNNIG 710
```

## RESULT 15

## VCPVCD

coat protein VP1 - canine parvovirus (strain CPV-d)

N:Contains: coat protein VP2

C:Species: canine parvovirus, CPV

C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999

C:Accession: A31163

R:Parish, C.R.; Aquadro, C.F.; Carmichael, L.E.

Virology 166, 293-307, 1988

A:Title: Canine host range and a specific epitope map along with variant sequences in th

A:Reference number: A31163; MUID:89020796; PMID:3176341

A:Accession: A31163

A:Molecule type: DNA

A:Residues: 1-737 <PAR>

A:Cross-references: EMBL:M23255; NID:g333467; PIDN:AAA47158.1; PID:g333468

## C:Genetics:

A:Introns: 26/3

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

F:584-737/Product: coat protein VP2 #status predicted <VP2>

```
Query Match      6.4%; Score 208; DB 1; Length 737;
Best Local Similarity 19.8%; Pred. No. 2.3e-06;
Matches 137; Conservative 96; Mismatches 260; Indels 200; Gaps 31;

QY  7 RPVEQSPQEPSSGIGKGTQQAQKRLNFGTGDSESV-----DPQLGEPPEA--TPAA 60
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 KPTKRSKPPPHIFINLAK-----KKKAGAGQVKRDNLAPMSDGAQVDPGGQPAVRNERA 173
QY  61 VQPTTMASGGAPMADNNEGADGVGNASGNWH-----CSTWLGDRVITSTRTWALP 113
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 TSGNGSGGG-----GGSGGVIGISTGFNNQTEPKFLENGWV--EITANSRLVHLN 225
QY  114 TYNHLYKQI-----SSASTGASNDNHFGYSTPMGYDFPNRHFCHFSPRDQRLINN 166
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 MPESYKRVVNNMDKTA VGNMALDDIHQIVTPWLSLVDANAWGVNFPNGDQWLIVNT 285
QY  167 NWGPRPKRLNPKLNIQVKEVT---TNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHQ 223
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 286 MSELHLVSFEQEIFNVVLKTVSESATQPTKYNNDLTASLWALDSSNTMPFTPAAMRS 345
QY  224 GCL-----PPPPA-----DVFMIPQGYLTLLNGSQAVGRS-----SFYCL 259
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 346 ETLGFPYKWTPTTPWRYFQWDRTLIPSH-----TGTSGTPTNVYHGTDPDDVQFYTI 399
QY  260 E-YFPSQMLRTGNF-TFSYTFEEVPHSSYAHQSQSLRLMN--PLIDQYLYLNRNQ 315
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 ENSVPVHLLRTGDEFATGTFDFCKP--CRLTHTWQTNRALGLP-----FLNSLPQS 450
QY  316 SGAQNKOLLPSRGSFAGMSVQPKNWLPGCYRQORVSKTKTDNNNSF-----TW 366
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 EGATNFGDI-----GV-----QDQKRGVQMGNTNYITEATMPAE 488
QY  367 TGASKYNLNGRESIINP-----GTAMASHKDEDEKFPFMSGVMIFKESAGASNTALD-- 419
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 VGSAPYSPSEASTQGFKTPIAAGRGGAQTDENQAAADGNPRYAFGRHQGKTKTTTGETP 548
QY  420 -----NYMTDEEIKATNPVATERFGTVAVNFSSSTDP 454
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 ERTYIAHQDTGRYPEGDWIQINFNLPVNDNVLLPTDPIG---GKTGINY--TNIFN 602
QY  455 ATGDVHANGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPMLMGFGFLK----- 504
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 603 TYGFLTALNNVP-----PVYPNGQIWDKEFTD-----LKPRLLHVNAPFV 642
QY  505 ---NPPQILIKNTFVPAN---PPAEFSATKPFASFIQYSTQSVSEIWEIWEIWEIWEIWEI 558
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 643 CONNCPQLFVKVAPNLNTEYDPPDASANMSR---IVTYSDFWPKWKGKLVFKAKLRASHTW 698
QY  559 NPEVQYTSNYAKSANVDFTVNNGLYTEPRPIG 591
DQ  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 NPIQMSIN-----VDNQFNIV--PSNIG 720
```

Search completed: January 21, 2004, 16:07:07

Job time : 16.1017 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	Length				
1	2177.5	67.0	504	1	COA3	AAV2	P03135 adeno-associ
2	556	17.1	1	COAT	PAVBO		P07297 bovine parvo
3	495	15.2	781	1	COAT	PAVHB	P07299 human parvo
4	254.5	7.8	729	1	COA1	PAVP9	P33484 porcine par
5	254.5	7.8	729	1	COA1	PAVPN	P18546 porcine par
6	252.5	7.8	729	1	COA1	PAVPK	P52501 porcine par
7	244	7.5	722	1	COAT	PAVHH	P03136 hamster par
8	234	7.2	587	1	COAT	PAVL3	P63310 parvovirus
9	216	6.6	716	1	COAT	MUMIV	P03137 murine minu
10	216	6.6	727	1	COAT	FPV	P04864 feline panl
11	214.5	6.6	718	1	COAT	MUMIM	P07302 murine minu
12	214	6.6	722	1	COAT	MEVA	P27437 mink enteri
13	209	6.4	727	1	COAT	FPV19	P24840 feline panl
14	208	6.4	727	1	COAT	PAVCB	P11213 canine parv
15	208	6.4	737	1	COAT	PAVCD	P17455 canine parv
16	204	6.3	748	1	COAT	PAVCN	P12930 canine parv
17	195	6.0	584	1	COAT	PAVC2	P30129 canine parv
18	193	5.9	722	1	COAT	PAVC7	P04863 canine parv
19	186.5	5.7	647	1	COAT	ADVG	P24029 aleutian mi
20	143.5	4.4	648	1	FXN1	MOUSE	Q61575 mus musculu
21	123	3.8	648	1	FXN1	HUMAN	O15353 homo sapien
22	121	3.8	1142	1	ENAM	PIG	O97939 sus scrofa
23	121	3.7	880	1	SVV	BAGST	P11931 bacillus st
24	121	3.7	1113	1	N116	YEAST	Q02630 saccharomyc
25	119	3.7	667	1	ITF2	HUMAN	P15884 homo sapien
26	116.5	3.6	642	1	ITF2	CANFA	P15881 canis fami1
27	113	3.5	781	1	SP3	HUMAN	Q02447 homo sapien
28	110.5	3.4	655	1	CDGT	KLEPN	P08704 klebsiella
29	110.5	3.4	1742	1	GUNA	CALSA	P22534 caldocellum
30	109.5	3.4	531	1	HEXB	PIG	Q29548 sus scrofa
31	109.5	3.4	670	1	ITF2	MOUSE	Q60722 mus musculu
32	109	3.4	1186	1	CEAA	BACTS	Q45710 bacillus th
33	108.5	3.3	860	1	AREA	PENRO	O13508 penicillium

Db 240 LYLSTNTPTSTQSRLOFQAGASDIRDQSRNWLPGCYRQVRKTSADNNSEYS 299  
 QY 366 WTGASKYNLNGRESINPTAMASHKDDKFFPMGSMVIFKESAGASNTALDNVMTD 425  
 Db 300 WTGATKYHLNGRSLVNP--ANASHKDDDEKFFQSGVLIFKQSGSEKTNVNIERKMTD 357  
 QY 426 BEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMVWQDRDVLQGGPIWAKI 485  
 Db 358 BEEIGTNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMVWQDRDVLQGGPIWAKI 417  
 QY 486 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAFSAATKFSFITYSTG 540  
 Db 418 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAFSAATKFSFITYSTG 472  
 RESULT 2  
 COAT\_PAVBO  
 ID COAT\_PAVBO STANDARD; PRT; 673 AA.  
 AC P07297; Q84374;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coat protein VP2 (Contains: Coat protein VP3).  
 OS Bovine parvovirus (BPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=10784;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87061184; PubMed=3783814;  
 RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,  
 Bates R.C.;  
 RT "Complete nucleotide sequence and genome organization of bovine  
 parvovirus";  
 RL J. Virol. 60:1085-1097 (1986).  
 CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M14363; AAB59848.1; -;  
 DR EMBL; M14363; AAB59849.1; -;  
 DR PIR; A26104; VCFVB5.  
 DR HSP; P30129; 4DPV.  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CHAIN 1 673 COAT PROTEIN VP2.  
 FT CHAIN 138 673 COAT PROTEIN VP3.  
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DOMAIN 163 187 GLY-RICH.  
 SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 17.1%; Score 556; DB 1; Length 673;  
 Best Local Similarity 25.0%; Pred. No. 2e-29;  
 Matches 160; Conservative 98; Mismatches 239; Indels 144; Gaps 19;

QY 1 TAPGKKRPVQSPOEISSGGTGTQQAQKELNFGTGDSESVDPQLGEP--PATP 58  
 Db 100 TSGGGRALKRKYFARSKGAKKAREPAPSTSNQONNEVNDIPNDEAGNQPIELATP 159  
 QY 59 AAVGPTTMASSGGAPMADNNEGADGVGNASGNHCDSTWLGRVITSTRTWALPTYNH 118  
 Db 160 SVVSGSVGGG-----RGSVGSGYGTGGTGTIFSENIVVTNTQFICDKNGH 211  
 QY 119 LYKQISSASTGASNDNHFGYSTPWGVDNFRNHFCHFSRDRWLRNNWGRPKRLNPK 178

Db 212 LYKS-EVLNTGDTAHOY-AITFWSYFNFOYSSHSPNDQHLVNDYERFRKAMIVR 269  
 QY 179 LFNIOQKVTNDGVTTIANN-LFSTVOVPSDSYQIPLYLGSAGHQCGLPPFPADVMP 237  
 Db 270 VYNLIQIKIMTDGAMGTVYNNDLTAGMHIFCGDHPHYVQHPWDDQCMPELPSIWELP 329  
 QY 238 QYGLTL-----NNGSOAVGRS-----SFYCLEVFPDSMLRTGNNFTFSYTPBEVPFHS 286  
 Db 330 QYATIPAPISVDNNTTVEEHLKGVPLYMENSDEVLNG----- 373  
 QY 287 SYAHSQSLDRMLNPLIDQYLYLNRQN-----QSGSAQNKKDLFSRGSP----- 331  
 Db 374 -----RIVRIVIQWLRLMRDKQHHIGHASDDVQSTGQKKNLLIQTOKPNKQRF 424  
 QY 332 AGMSVQPKNWLPGCYRQVRKTSKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHK 391  
 Db 425 QNAALRTSNMWSGP-----GIARTHTNATLQTQAGALVTWVT 462  
 QY 392 DEEDKFFPMGSMV-----IFGKE-----SAGASNTALDNVMTDE 426  
 Db 463 NGAD-----VSGVRAVRVGYSTDPYGGQOPESDILLRLYSASAAEGQONPILEN----- 512  
 QY 427 BEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMVWQDRDVLQGGPIWAKIP 486  
 Db 513 -----AARHTFTREATKLTGSGNAGDGDYKSWMLPNQNMDSAPISRYNPPIWVKVP 564  
 QY 487 HTDGHFHPSPLMGGFGLKNPPQILIK--NTPVPANPPAFSAATKFSFITYSTGQSV 544  
 Db 565 RVNRKTLDTQDGSIPMSHPPTGTFIKLARIPVPGND-----SFLNIYVTGQVSC 615  
 QY 545 EIEWELQKSKRNWNPVQYTSNYAKSANVD--FTVDNNGLY 584  
 Db 616 EWWVEVERKGTKNWRPEYMHs---ATNMSVDAYTINNAGVY 653  
 RESULT 3  
 COAT\_PAVHB  
 ID COAT\_PAVHB STANDARD; PRT; 781 AA.  
 AC P07299;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 01-APR-1990 (Rel. 14, Last annotation update)  
 DE Probable coat protein VP1.  
 OS Human parvovirus B19.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.  
 OX NCBI\_TaxID=10798;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Isolate AU;  
 RX MEDLINE=86200451; PubMed=3701931;  
 RA Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;  
 RT "Nucleotide sequence and genome organization of human parvovirus B19  
 isolated from the serum of a child during aplastic crisis";  
 RL J. Virol. 58:921-936 (1986).  
 CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M13178; AAB6867.1; -;  
 DR PIR; A24299; VCPV19.  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 293 293 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 781 AA; 86015 MW; 8C6254DBD0576B07 CRC64;

Query Match  
Best Local Similarity 27.3%; Score 495; DB 1; Length 781;  
Matches 148; Conservative 84; Mismatches 257; Indels 54; Gaps 16;

QY 51 LQEPATPAAGPTTASGGGAPMADNNEGADGVGNASGWHCDSTWLDGRVITTSRTW 110  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 213 LPEVPAYNASEKYPMTSVNSAE-ASTGAGGCGSNVSKSMWEGATFSANSTCTFSRQF 271  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 111 ALPTNNHLYKQISSASTGASND-----NHYFGYSTGPGVDFNRFCHSPRDW 160  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 272 LIPYDEHYKVFSPAASSCHNASGKAKVCTISPMGYSTPWRVLDNFALNLFSPLEF 331  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 161 QRLINNMGFRKRLNFKLFNIQVKEVT--TNDGVTTIANNTSTVQVPSDEYQLPYVL 218  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 332 QHLIENYGSIAFDALVTITSEAVKDVTKTGGV-QVTDSTIGRLCMLVDHEYKYPYL 390  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 219 GSAHQCLPPFPADVPMIPQGYLTUNN-GSQAVG-----RSSFYCLFYFSPQMLR 268  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 391 GQGQDTLAPLPIWVFPQYAYLTVDVNTQGISGDSKKLASSESAFYVLESHSGFQLLG 450  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 269 TGNFTFSTVEEVPFHSSVAHSOSLDRLNPLIDQYLYLHRTQNSGSAQNKOLLFSR 328  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 451 TGTASMSYKPPVPENLEGCSQHPYEMNPL---YGRLVGPDVTLGGDPKFRSL---- 503  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 329 GSPAGMSVQPKWLPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINFG-TAM 387  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 504 -THEDHAIQONFMFGPLNSVSTKEGDSNTGAGKALTGLSTGTSQNTSLRPGPVSQ 562  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 388 ASHKDEDEKFPMSGVMIFGKESAGASNTALDNV-----MITDEEKATNPVATERF 440  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 563 PYHWDTKYVTGAINAISHGQTTYG---NAEDKEYQQGVGRPNKEQQLQGLNMEHY 619  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 441 GTVAVNFQSSSTDPAITGVHMGALPGMWQDRDVTYVLOGPIWAKIPHTDGHFHS-PLMG 499  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 620 -----PFNKGQTYTDQIE-RPLMVGSVNRRALHVESQLSKIPNLDSPKTFQFALG 672  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 500 GFLKNPPQILIKNTVPANPAEFSATKFSFITQSTGVSVIEIWEIWEIWEIWEIWEI 558  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 673 GWGLHQPPIELK--ILPQSGPIGGIKSMGITTLLVQYAVGIMTWTMTFKLPGKATGRW 730  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 559 NPE 561  
Db CARBOHYD 273 273 N-LINKED (GLCNAC... ) (POTENTIAL).  
QY 731 NPQ 733

RESULT 4  
COAL\_PAVP9 STANDARD; PRT; 729 AA.  
AC P33484;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Porcine parvovirus (strain 90HS) (ppv).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=33725;  
RN [1]  
RX MEDLINE=89319168; PubMed=2750278;  
RA Sakurai M., Nishimori T., Uehimi C., Nakajima H.;  
RT "Nucleotide sequence of capsid protein gene of porcine parvovirus.";  
RL Virus Res. 13:79-86(1989).  
CC -1- SUBUNIT: NATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
CC VPI, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.  
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
DR PIR; A60006; A60006.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.

DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 729  
FT CHAIN 151 729 COAT PROTEIN VP1.  
FT CARBOHYD 172 172 COAT PROTEIN VP2.  
FT CARBOHYD 198 198 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 282 282 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 433 433 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 471 471 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 573 573 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 604 604 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 651 651 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 729 AA; 80938 MW; 92538BFF9A0C78E6 CRC64;

Query Match  
Best Local Similarity 22.7%; Score 254.5; DB 1; Length 729;  
Matches 151; Conservative 95; Mismatches 277; Indels 141; Gaps 33;

QY 9 VEQSPQEPDSSSGIGK-----TQQQPAKRLNFGQTDSESVDPDQPLGEPATP 58  
Db VRSRPRKHGSKPPGKRPAPRHIFINLAKKAKGTSNTNSNSMENVEQHPINAGTEL- 167  
QY 59 AAVGPTTASGGAPMADNNEGADGVGNASGWH--CDSTWLD--RVITTSRTWALP 113  
Db SATGNESGGGGGG---GGRGAGGVSTGSGFNQTEFOYLGELVRAITAHASRLHLN 223  
QY 114 TYNHLYKQIS--SASTGAS---NDNHYFGYSTGPGVDFNRFCHSPRDWQRLINN 167  
Db MPEHETKRIHVLNSESAGVQVQDDAHTQVTFWLSLDANAGWFWNPADWLISNM 283  
QY 168 WGRPRKLNFLKFNQVKEVT--TNDGVTTIANNTSTVQVPSDEYQLPYVLSAHQ 224  
Db TEINLVSEQEIFNVLKTIETESATSPETKIYNDLTASLMAVALDTNNTLPTPAARSE 343  
QY 225 CLPFPADVPMIPQGYL-----TLNNGSQVGRS-----SYCLE-YFPS 264  
Db TLGYFPLPTKPTQYRYLSCTRNLPPTYTGOSQITDSIQTLGHSIDIMFYTIENAVPI 403  
QY 265 QMLRTGNF-TFSYTFEEVPHSSVAHSQSILDR-----LMNPLI--DOVLYYLNRQ 313  
Db HLLATGDEFFSIGIYHFDTKPL--KLTHSWQTNRSISGLPPKLLTEFTTEGDHPTLPAAN 461  
QY 314 NQSGAQNKLFLSRGSPAGMSVQPKWLPGPCYRQORVSKTKTDNNNSNFTWTGASKYN 373  
Db TRKGVHTMNSYTEAT-----AIRP-----AQVGYNTPYMNPEYSGGPP- 502  
QY 374 LNGRESIINP--GTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDNVITDEEIK 431  
Db LTPIVPTADTYNDDE---PNGAIRFTMGYQHGLTTS-----SOELERYT 545  
QY 432 TNP-----VATERFGTVA-VNFQSS-----TDPATG--DVHAMGALP-----GMVWQ 471  
Db FNPOKCGRAPKQPNQAPLNTNNGTLLPSDPIGCKENMFMTLNTYGPPLTALNN 605  
QY 472 DRDVLQGPWAKIPHTD--GHFHPSPLMGFGFLK-NPPQILIKNTVPANPAEFS- 527  
Db TAPVFPNGQIWDKELDTLKLRLH---VTAPFVCKNNPPGQLFVKIAD---NLTDNFAD 659  
QY 528 TKFASFTQYSTGVSVIEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEI 587  
Db SPOQPRIITYSNFWKGLTFTAKRSSNMNPIQQHTT-----TAENIGNYI-P 708

RESULT 5  
COAL\_PAVPN STANDARD; PRT; 729 AA.  
ID COAL\_PAVPN  
AC P18546; P22964; Q89816;  
DT 01-NOV-1990 (Rel. 16, Created)





QY 159 DMQRLNNMGRPKLNFKNLQVKEVTT-----NDGVTTIANNTLSTVQVSDSEYQ 213  
 DB 259 DMQFIQNSMESLNDLSQLEFNVVVKTVEQOAGQADAKVYNNDLTACMMVALDSNNI 318  
 QY 214 LPVVLGSAHQGL-----PPFPAD---VFMIPOGYLTLLNGSQ-----AVG-----R 253  
 DB 319 LPVTPAAQTSITLGFYFWKPTAPAPRYFFPMPRQLSVTSSSAEQTQITDIFGPAQLN 378  
 QY 254 SGPYCLE-VFPPSOMLRTGNFTF-SYTFEEVPHSSYA-----HSQSLDLAMN-PLIDQY 305  
 DB 379 SQPFTTIENTLPTLLRIGDEFTGTGTFYNTDPLKLTHTWQNRHLACLOGITDLPSTDA 438  
 QY 306 LYYLNTQNSQSAQNKDLF-----SRGSPAGMSVQPKWL 342  
 DB 439 TASLTANGDRFGSTQTONVYVTEALTRPAQIGFMQPHDNFEANRGGPKFVPVVP----- 494  
 QY 343 PGPCYQQRVSKTKTDNNNSNFTWTSKYNLNGR---ESLIIPGTA-----MASH 390  
 DB 495 -----LDITAGEDHDAN-----GAIRFNYGKHGEDWAKQGAAPERITWDALDSAG 541  
 QY 391 KODEDKFFPMGSMVIMFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQS 449  
 DB 542 ROTARCFV-----QSAPISIPNQNILOREDALAGRTNMHYTNVNSYGPPLSAF 591  
 QY 450 SSTDPATGADVHANGALPGMWQDRDYVLOGPIWAK---IPHTDGHFHPPLMGGFGLKNP 506  
 DB 592 PHDP-----IYPNGQIWDKELDEHKPLHVTAPFV---CKNNP 628  
 QY 507 PPQILKNTVPANPAEP---SATKASFTQYSTQVSVIEIWELOKENSKEWNPVQY 564  
 DB 629 PGOLFVHLGP---NLTDQDPDNTTSVRIVT-YSTFYWKGLFKAKLPNLTWNPVYQA 684  
 QY 565 TSN 567  
 DB 685 TTD 687  
 RESULT 8  
 COAT\_PAVL3 STANDARD; PRT; 587 AA.  
 AC P36310;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coat protein VP1 [Contains: Coat protein VP2].  
 OS Parvovirus Lu111.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=35339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93297126; PubMed=8517025;  
 RA Diffot N., Chen K.C., Bates R.C., Lederer M.;  
 RT "The complete nucleotide sequence of parvovirus Lu111 and  
 RT localization of a unique sequence possibly responsible for its  
 RT encapsidation pattern.";  
 RL Virology 192:339-345(1993).  
 CC -/- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC ENBL; M81888; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; B44276; B44276.  
 DR HSSP; P07302; 1MWV.  
 DR InterPro; IPR001403; Parvo coat.  
 DR Pfam; PF00740; Parvo coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 587 AA; 65429 MW; 523E6B9CBF2BB74 CRC64;  
 Query Match 7.2%; Score 234; DB 1; Length 587;  
 Best Local Similarity 23.1%; Pred. No. 3.4e-08;  
 Matches 137; Conservative 74; Mismatches 231; Indels 152; Gaps 29;  
 QY 59 AAVGPTTASGAGPADNNEGADGVGNASGNWHCHDS--TWLGD---RVITSTRTWAL- 112  
 DB 23 AADGPG--GSGGG-----GSGGGGVGVSTGSYDQTHYKFLGDGWVEITAYSTRVHLN 74  
 QY 113 -PTYNN-----HLYKQISSASTGASNDNHYFGYSTPMGYDFNRFCHFSRPRQRLIN 165  
 DB 75 MPKSENYCRVRVHTNDTGTASHMAMDDAHEQIW-TPWSLVDANAWGVWFPQSDQYISN 133  
 QY 166 NNWGRPRKRLNFKLNIQVKEVT-TNDGVTTIA-----NNLTSTVQVFSSEYQLPYVLSA 221  
 DB 134 NMHINLHSLDQELFNVVIKTVTEQNTGAEAKVYNNDLTAAMVVALDSNNILPYTPAID 193  
 QY 222 HQGCL-----PPPPA-----DVFMIPOGYLTLLNGSQAVGRSFYCLEYFP 263  
 DB 194 NQETLGFPWKEPTIPSPRYFYFCDRNLSTVYKDEAGTITDMGLASGLNSQFFTIENQ 253  
 QY 264 S-QMLRTGNF--TFSYTFEEVPHSSYAHOSLSRLMPLIDQ-----YL 306  
 DB 254 RINLLRTGDEVATGYTFDTEPIRLTHTWQNRHLGQPPQITELPSSDTANATLTARGYR 313  
 QY 307 YLNTQNTON-----QSGSAQNKD--LLFSRGSAGMSVQPKWLPG----- 344  
 DB 314 SGLTQIGRNDVTEATRVPAQVGCQPHDNFETSARGFFKVPVVPADITQGLDHDANGS 373  
 QY 345 --PCYRQQRVSKTKTDNNNSNFTWTSKYNLNGRESIINPGTAMASHKDDKDFPMSG 402  
 DB 374 LRYTYDKQHGQSWASQNNKDRYTW-DAVNYD-SGR----- 406  
 QY 403 VMIFGKESAGASNTALDNVMTDDEEIKATNPVATER---FGTVAVNFQS---SSTDPA 457  
 DB 407 -----WTNNCFIQSVPTSEPN---ANQILTRNDNLAKGTIDHFTNAFNSYGLPLTA 454  
 QY 458 DVHANGALPGMWQDRDYVLOGPIWAK---IPHTDGHFHPPLMGGFGLKNPQILIKN 514  
 DB 455 FPH-----PAPIYQGGIWDKELDEHKPLHVTQAPFV---CKNNAPGQLLVRL 500  
 QY 515 TP--VPANPPAFSFKPASFTQYSTQVSVIEIWELOKENSKEWNPVQYTS 566  
 DB 501 APNLTQDQDPNSSLNLSRVITYGTFFPKGKLTILKAK---MRPNA-TWNPVYQISA 550  
 RESULT 9  
 COAT\_MUMIV STANDARD; PRT; 716 AA.  
 AC P03137;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coat protein VP1 [Contains: Coat protein VP2].  
 OS Murine minute virus (Murine parvovirus).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=10794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83143341; PubMed=6298737;  
 RA Astell C.R., Thomson M., Merchlinsky M., Ward D.C.;  
 RT "The complete DNA sequence of minute virus of mice, an autonomous  
 RT parvovirus.";



Nucleic Acids Res. 11:999-1018(1983).

CC -I- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; V01115; CAA24310.1; ALT\_SEQ.

DR PIR; A03700; VCPV2M.

DR HSSP; P07302; 1MVM.

DR InterPro; IPR001403; Parvo\_coat.

DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein.

FT CHAIN 1 716 COAT PROTEIN VP1.

FT CHAIN 131 716 COAT PROTEIN VP2.

FT CARBOHYD 179 179 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 669 669 N-LINKED (GLCNAC. .) (POTENTIAL).

FT DOMAIN 156 171 GLY-RICH.

FT SEQUENCE 716 AA; 78707 MW; 6A7229A91161F4C6 CRC64;

Query Match 6.6%; Score 216; DB 1; Length 716;

Best Local Similarity 20.8%; Pred. No. 6.8e-07;

Matches 148; Conservative 93; Mismatches 271; Indels 198; Gaps 34;

QY 8 PVEQSPQEPDSSGIGKTG-----QPAKRL-----NFGQT-GDSESVDP 48

DB 82 PKLATSEP-GTSGVSRAGKTRPPAYIFINQARAKKLTSSAAQSSQTMDSGTSDPS 140

QY 49 QPLGEPPA--TPAUGPTTWSGGAPMADNNEGADVGNAGNWHCD--TWLGP---R 101

DB 141 GNAVHSAARVERAADGPG--GSGGG-----GSGGGVGSTGSDYNQTHRYFLGDGWYE 192

QY 102 VITSTRTRWALPTYNHLYKQI-----SSASTGASNDNHVFGYSTPMGYDFENRFCH 154

DB 193 ITALATRLVHLNPKSENCRTRVHNTTTSVKNWAKDAHQIWTWPSLVDAWAGVW 252

QY 155 FSPRDWRLNNWGRPRKLNKFLNIOVKEVTND-----GVTTIANLSTVQVFS 210

DB 253 LQPSDWQVICTMSQLNLVSLDQIEFNVLKTVTEQDLGQAIKIYNNDLTACMWAVDS 312

QY 211 EYQPLVLSAGHQCGLPPPADVFMIPQYGY-----LTLNQSQAV-----G 252

DB 313 NNILPTTPAANSMETLGFYFWKPTIASPYRYFCVDRDLSTVYENQEGTVHNVMTGPKG 372

QY 253 RRSFYCLEYFPS-QMLRTGNF--TFSYTFEEVPFHSYAHSSQSLRLMN--PLIDQLYY 308

DB 373 IPOFFTIENTQITLRTGDEATGYFDTSV--KLTHWTQNRQLQPPLLSTP--- 427

QY 309 LNRNQSSGAQNKOLLFRGSGPAGMSVQPKWL-----PG 344

DB 428 -PEADTDAGT-----LTAQGRHGTTQMGVNVSEAIRTPAQVGCQPHNDPEASRAG 480

QY 345 PCYQROVSKTKTDNNNSFTWTGASKYNLNGRESIINPG-----TAWASHKODE 394

DB 481 P-FAAPKVPADITQGVDKBANGSVRYSGKQGENWASHGAPERYTWDTSFGSGRDTK 539

QY 395 DKFFP-----PMSGVMIKFESAGASNTALDNNWITDEBEIKATNPVATERFCTVAV 445

DB 540 DGIQAPLVVPPPLNGI-----LTWNPITGN-----DI 570

QY 446 NFQS---SSTDPTAGDVHANGALPGMWQDRDQVYVQPIWAK-----IPHTDGHFHPSPMGG 500

DB 571 HFSNVFNSYGLTAFSH-----PSPVYQGIWDKELDLKHLRLHITAPFV-- 617

QY 501 FGLKNPQPILLIKTP--VPANPPAFSATKFSPTQYSTGVSVIEIWELOKENSKEW 558

DB 618 -CKNNAPQMLVRLGPNLTDQDPNGATLSRIVYTYGTFPWKGLTMRKLA---NTTW 672

QY 559 NPEVQYTSNYAKSANDVTVNNGLY---TEPRPIGT-----RYLTPRL 599

DB 673 NPEVQVSAE-----DNGNSYMSVTKWLPTATGNMQSVPLITRPV 711

RESULT 10

COAT\_FPV STANDARD; PRT; 727 AA.

AC P04864; Q65112;

DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

OS Coat protein VP1 [Contains: Coat protein VP2].

OS Feline panleukopenia virus (FPV)

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.

OX NCBI\_TaxID=10786;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85265017; PubMed=2991581;

RA Carlson J., Rushlow K., Maxwell I., Maxwell F., Winston S., Hahn W.;

RT "Cloning and sequence of DNA encoding structural proteins of the

RT autonomous parvovirus feline panleukopenia virus.";

RL J. Virol. 55:574-587(1985).

CC -I- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M10824; AAA47161.1; -

DR PIR; A03701; VCPVIF.

DR HSSP; P30129; 4DPV.

DR InterPro; IPR001403; Parvo\_coat.

DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein.

FT CHAIN 1 727 COAT PROTEIN VP1.

FT CHAIN 144 727 COAT PROTEIN VP2.

FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 207 207 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 323 323 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 586 586 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. .) (POTENTIAL).

FT DOMAIN 165 180 GLY-RICH.

FT SEQUENCE 727 AA; 80344 MW; 77E0F6FES54C0C6E CRC64;

Query Match 6.6%; Score 216; DB 1; Length 727;

Best Local Similarity 20.2%; Pred. No. 6.9e-07;

Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;

QY 7 RPVEQSPQEPDSSGIGKTGQPAKRLNFGQTDGSESV-----DPOPLGEPPA--TPAA 60

DB 110 KPTKRSKPPPHIFINLAK-----KKKAGAGOVKRDNAQPMDSGAVQPDGGQFVNRBA 163

QY 61 VGPTTWSAGGGAPMADNNEGADVGNASGNWH-----CDSTWLGDRVITTTSTRTWALP 113

DB 164 TSGNGSGGG-----GGSGGVGISTGTNNQTEPKFLENGWV--BITANSSRLVHLN 215

QY 114 TYNHLYKQI-----SSASTGASNDNHVFGYSTPMGYDFENRFCHFSRDLWRLINN 166

DB 216 MPSESNYKRVVNNMDKTAVKGNMALDDTHVQIVTPWLSLVDAWAGVWFPDGMQLIVNT 275

QY 167 NWGFPRKLNKFLNIOVKEVT---TNDGVTTIANLSTVQVFSDESQVLPYVLSAHQ 223

DB 276 MSELHLVSFBOEIFNVVLKTVSESATQPTTKYNNDLTASLVALDSNNTMFTTAAAMRS 335



```
FT STRAND 652 668
FT STRAND 683 683
FT STRAND 691 691
FT HELIX 692 694
SQ SEQUENCE 718 AA; B43C1762ED5F74B9 CRC64;

Query Match
Best Local Similarity 21.0%; Score 214.5; DB 1; Length 718;
Matches 142; Conservative 89; Mismatches 273; Indels 171; Gaps 30;

QY 8 PVEQSPQBPSSGIGKGTG-----QQPAKKRLNFG---QTGDSSEVPDPQPLG 52
Db 83 PGLATUSEP-GRGVRAGKRTPPAYIINQRAKKLUTSSAAQSSQSDGTSQPQD 141
QY 53 -----EPPATPAAGVPTTMASSGGGAPMADNNEGADGVGNAGNHCDS--TWLGD---R 101
Db 142 GNAVSAARVERAADGFG--GSGCG-----GSGGGGVSTGSDYNQTHYRFLGDGWE 193
QY 102 VITTTTRTVALPTNNHLYKQI-----SSASTGASNDNHFGYSTPNGYDFDNRPHCH 154
Db 194 ITALATRLVHLNMPKSENCRIRVHNTTDTSVKGNMAKDDAHEQIWTWPSLVDANAAGVM 253
QY 155 FSPDRQRLNNWGRPKRLNPKLFNIQVKEVTND-----GVTTIANNLTSVQVFS 210
Db 254 LQPSDQYICNTMSQLNLSLQDEIFNVVLKTVTEQDSGQAIKIYNDLTACMVAVDS 313
QY 211 EYOLPYVLGSAHGGCLPPPPADVFMIPOXY-----LTLNN-----GSQAV 251
Db 314 NNILPYTPAANSMETLGFYFWKPTIASPYRYFCVDRDLSVTYENQEGTIEHNVMTGPK 373
QY 252 GRSSPYCLEYFPPS-QMLRTGNRP-TFSYTFEEVFPFHSVAHSQSLRLMN--PLIDQYLY 307
Db 374 MNSQPFTEIENQOITLLRTGDEPATGYTFDTPNV--KLTHWTQTNRLQGPPLSTF-- 429
QY 308 YLNRTQNGSAGNKKLLSRGSPAGMSVQPKWL-----P 343
Db 430 --PEADTDAGT-----LTAQSRHGATQMEVNVWSEAIRTPAQVGFQCPHNDPEASRA 481
QY 344 GPCYRQORVSKTKTDNNNFTWTGASKYNLNGRESIINPGTAMASHKDEDKP----- 397
Db 482 GP-FAAPKVPADVTQGVREANGSVRYSGKHGENWAAHGPAPERYTWDETNFGSGRDT 540
QY 398 ---PPMSGVMI FCKESAGASNTALDNVMTDDEEIKATNPVATERPGTVAVNFQS--SST 452
Db 541 RDGFIOQAPLWPPPLNGLITNA-----NPIGTQN-----DIHFSNVFNSY 581
QY 453 DPATGDVHAMGALPGMWQDRDVLVLOGPIWAK---IPHTDGHFHPSPMLGGFGLKNPPQ 509
Db 582 GLPTAFSH-----PSPVYFQQLWDXELDLEHPRHLHITAPFV---CKNAPQ 627
QY 510 ILIKNTP--VPANPPAFSATKPAFTQYTSQGVSEIWELOKENSKRNPVQYTSN 567
Db 628 MLVRLGPLNTDQVDPNGATLSRIVTYGTFFWKGKLTMRALRA---NTTNVPVQ---- 679
QY 568 YAKSANVDFTVNDNG 582
Db 680 -----VSVDNG 686

RESULT 12
COAT_MEVA
ID COAT_MEVA STANDARD; PRT; 722 AA.
AC P27437;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coate protein VP1 [Contains: Coat protein VP2].
OS Mink enteritis virus (strain Abashiri) (MEV).
OC Viruses; ssRNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID:10793;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91202123; PubMed-2016597;
```

```

QY 461 AMGALPGVMWQDRVYVLOGPIWAKIPIHTDGHFHPSPLMGGFLK-----NPP 507
Db 594 ALNNVP-----PYPNGQIWDKEFTD-----LKPRLVHNAFFVCQNCNP 633
QY 508 POLIKNTVPVPA-----PPAEFSATKFAPIFYTOYSGVSVIEWELQKENSQRNNEVQY 564
Db 634 GQLFKVAPNLNTEYDPPASANMSR-----IVTYSDFWMKGLVFKAKLRASHTNWPIQOM 689
QY 565 TSNTAKSANVDFTVDNNGLYTEPRPIG 591
Db 690 SIN-----VDNQFNYL--PNNIG 705

RESULT 13
COAT_FPV19
ID COAT_FPV19 STANDARD; PRT; 727 AA.
AC P2490;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Feline panleukopenia virus (strain 193) (FPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=193/70;
RX MEDLINE=91073139; PubMed=2174965;
RA Martyn J.C., Davidson B.E., Studdert M.J.;
RT "Nucleotide sequence of feline panleukopenia virus: comparison with
RL canine parvovirus identifies host-specific differences.";
RJ J. Gen. Virol. 71:2747-2753(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CU-4;
RX MEDLINE=91272479; PubMed=1647068;
RA Parrish C.R.;
RT "Mapping specific functions in the capsid structure of canine
RT parvovirus and feline panleukopenia virus using infectious
RT clones.";
RL Virology 183:195-205(1991).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55115; CAA38911.1; -
DR EMBL; M38246; AAC37928.1; -
DR EMBL; M38245; AAC37929.1; -
DR FIR; B36608; VCFVPF.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 727 COAT PROTEIN VP1.
FT CHAIN 144 727 COAT PROTEIN VP2.
FT DOMAIN 165 190 GLY-RICH.
FT CARBOHYD 168 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 323 323 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 660 660 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DOMAIN 165 180 GLY-RICH.
SQ SEQUENCE 727 AA; 80386 MW; 648596C09B621FF5 CRC64;
Query Match 6.4%; Score 209; DB 1; Length 727;

```

```

Best Local Similarity 20.1%; Pred. No. 2e-06;
Matches 138; Conservative 101; Mismatches 260; Indels 188; Gaps 34;

QY 7 RPVEQSQEPDSSGIGKTKQPAKKELNFGTGDSESV-----DPOPLGEPPA--TPAA 60
Db 110 KPTKRSKPPPHIFINLAK-----KKAGAGQVKRONLAPMSGAVQPDGQQPAVRERA 163
QY 61 VGPFTMASGGGAPMADNNEGADGVGNASGNWH-----CDSTWLGDRVITTTTRTWALP 113
Db 164 TSGCNGSGGG-----GGSGGVGISTGTNNCTEFKFLENGWV--EITANSRLVHLN 215
QY 114 TYNHLYKQI-----SSASTGASNDNHYFGYSTPWGYDFDRFCHFPSPDRQORLINN 166
Db 216 MPESENTKRVVVMNDKTAVKGNMALDDIHQVITPWSLVDANAWGVWFNGDQQLIVNT 275
QY 167 NWGFRPKLNFKNIOVKET---TNDGVTTIANLLTSTVQVSDSEYQLPYVLGSAHQ 223
Db 276 MSELHLVSFEQIEFNVVLKTVSESATQPPKYVNDLTASLWVALDNNNTPTFAAMRS 335
QY 224 GCL-----PPFPA-----DVFMIPQYGYLTLNNGSQAVGRS-----SFYCL 259
Db 336 ETLGFYPMKPTIPTPMRYFQWDRTLIPSH-----TGTSGPTNVYHGTDPDDVQFYTI 389
QY 260 E-YFPSQMLRTGNMF-TFSYTFEEVPPHSSVAHSQSLDRLMN--PLIDQILYLYLNRQNO 315
Db 390 ENSVPVHLLRTGDEFATGTFPFCCKP--CRLTHTWQTNRALGLPP-----FLNSLPQS 440
QY 316 SGSQAKNDLLFSRSGSPAGMS-----VQPKN-WLPGCYRQORVSK-----354
Db 441 EGATNFGDIGVQOKRRGVQMGNTDYITEATIMRPAEVGSAPYYSFEASTQGPFTPI 500
QY 355 -----TKTDNNNSNFTWTGASKY---NLNGRESIINPGT---AMASHKDEDDKFPFMS 401
Db 501 AAGRGGAQTDENQAA---DGDPRYAFGRQHGQKTTTGTETPERFTYIAHQDT-----549
QY 402 GVMIFGKESAGASNTALD-NVMITDEBEIKATNPATERFCTVAVNFQSSSTDATGDVH 460
Db 550 -----GRYPEGDWIONFNLFTVNDVLLPTDPDG---OKTGINY--TNIFNTYGLPT 598
QY 461 AMGALPGVMWQDRVYVLOGPIWAKIPIHTDGHFHPSPLMGGFLK-----NPP 507
Db 599 ALNNVP-----PYPNGQIWDKEFTD-----LKPRLVHNAFFVCQNCNP 638
QY 508 POLIKNTVPVPA-----PPAEFSATKFAPIFYTOYSGVSVIEWELQKENSQRNNEVQY 564
Db 639 GQLFKVAPNLNTEYDPPASANMSR-----IVTYSDFWMKGLVFKAKLRASHTNWPIQOM 694
QY 565 TSNTAKSANVDFTVDNNGLYTEPRPIG 591
Db 695 SIN-----VDNQFNYV--PNNIG 710

RESULT 14
COAT_PAVCB
ID COAT_PAVCB STANDARD; PRT; 727 AA.
AC Q11213;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Canine parvovirus (strain B) (CPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=59284;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91272479; PubMed=1647068;
RA Parrish C.R.;
RT "Mapping specific functions in the capsid structure of canine
RT parvovirus and feline panleukopenia virus using infectious
RT clones.";
RL Virology 183:195-205(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC OF A COMBINATION OF VP2, VP3, AND SOME VP1.

```

CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M38245; AAB02799.1; -;  
DR EMBL; M38245; AAB02800.1; -;  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 727 COAT PROTEIN VP1.  
FT CHAIN 144 727 COAT PROTEIN VP2.  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 207 207 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 323 323 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 586 586 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 660 660 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 727 AA; 80342 MW; 115F3E1A79098EBE CRC64;  
  
Query Match 6.4%; Score 208; DB 1; Length 727;  
Best Local Similarity 19.8%; Pred. No. 2.3e-06;  
Matches 137; Conservative 96; Mismatches 260; Indels 200; Gaps 31;  
  
QY 7 RPVEQSPQBPSSSGIGKGTQGPAPKRLNFGQTGDSSEVP-----DPQPLGEPPA--TPAA 60  
DB 110 KPTKRSKPPPHIFINLAK-----KKKAGAGQVQRDLNAPMSDGAQVQDGGQPAVRNERA 163  
QY 61 VGPPTWASGGGAPMADNNEGADGVGNASGNWH-----CDSTWLGDRVITTTSTWALP 113  
DB 164 TSGNGSGGGG-----GGSGGGVGISTGTFTNNQTEFKLENGWV--EITANSRLVHLN 215  
QY 114 TYNHLYKQI-----SSASTGASNDNHFGYSTPMGYFEENRPHCHFSPRDWORLINN 166  
DB 216 MPESYRVRVNNMDKTAANGNMLADDDIHAQIVTPMSLVANAGWVFNPGDWQLIVNT 275  
QY 167 NMGRFRKRLNFKLFIQVKEVT---TNDGVTTIANNLSTTVQVFSDSYQLPVLGSAHQ 223  
DB 276 MSELHLVSPQEQEIFNVVLKTVSSAQTPTKVYNNDLTASLVALDSNNTMFTPAAMES 335  
QY 224 GCL-----PPPPA-----DVFMIPOGYLTLNAGSQAVERS-----SFYCL 259  
DB 336 ETLGLFYEWKPTIPTWRYYPQMDRTLIPSH-----TGTSGTPTNIYHGTDDVDQFYTI 389  
QY 260 E-YFPQMLRTGNF-TFSTVTFEEVPHSSYAHQSQSLRLMN--PLIDQVLYLNRNTQ 315  
DB 390 ENSVPVHLRTGDEFATGTFPFDCKP--CRLTHWTQTNALGLPP-----FLNSLPQS 440  
QY 316 SGSAQNKOLLFSKSGPAGMSVQPKWLPGPCYRQRVSKTKTDNNNSNF-----TW 366  
DB 441 EGATNFGDI-----GV-----QDQKRGVQMGNTNYITEATIMPAE 478  
QY 367 TGASKYNLNGRESIINP-----GTAMASHKDEDEKFFPMGSGVIMFKESAGASNTALD-- 419  
DB 479 VGSAPYSPFEASTQGFKPTIAAGRGGAQTDENQAADGNPRYAFQRHQKTTTGTETP 538  
QY 420 -----NMVITDDEIKATNPVATEREGTVAVNFQSSSTD 454  
DB 539 ERFTYIAHQDTGRYPGDMWIONINFLPVTDNVLPLTDPDG-----GKTGINV--TNIN 592  
QY 455 ATGDVHAMGALPGMWQDRDVLQGPPIWAKIPIHTDGHFHPSPLMGGFGLK----- 504  
DB 593 TVGPLTALNNVP-----PVYPNGQIWDKEFTD-----LKPRLHYNAPV 632  
QY 505 ---NPPPOLIKKTPVPAN---PPAEFSATKFASTIQTSTQVSVIEIWELOKENSRRW 558

DB 633 QNNCPGOLFVKVAPNLNLTNEYPDASANMSR-----IVTYSDFWMKGLVFKAKLRASHTW 688  
QY 559 NPVQVYTSNYAKSANDVFTVNNGLYTEPRPIG 591  
DB 689 NPIQMSIN-----VDNQFNIV-PSNIG 710  
  
RESULT 15  
COAT\_PAVCD STANDARD; PRT; 737 AA.  
ID COAT\_PAVCD  
AC P17455;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2]  
OS Canine parvovirus (strain CPV-D Cornell 320) (CPV).  
OC Viruses; sedna viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10790;  
RN [1]  
RP MEDLINE=89020796; PubMed=3176341;  
RX Parrish C.R., Aquadro C.F., Carmichael L.E.;  
RA "Canine host range and a specific epitope map along with variant  
RT sequences in the capsid protein gene of canine parvovirus and related  
RT feline, mink, and raccoon parvoviruses.";  
RL Virology 166:293-307(1988).  
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC OF A COMBINATION OF VP2, VP3, AND SOME VP1.  
CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M23255; AAA47158.1; -;  
DR PIR; A31163; VCPVCD.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 737 COAT PROTEIN VP1.  
FT CHAIN 154 737 COAT PROTEIN VP2.  
FT CARBOHYD 178 178 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 596 596 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 658 658 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT DOMAIN 175 192 GLY-RICH.  
SQ SEQUENCE 737 AA; 81504 MW; 39FE81D3E5435EC9 CRC64;  
  
Query Match 6.4%; Score 208; DB 1; Length 737;  
Best Local Similarity 19.8%; Pred. No. 2.4e-06;  
Matches 137; Conservative 96; Mismatches 260; Indels 200; Gaps 31;  
  
QY 7 RPVEQSPQBPSSSGIGKGTQGPAPKRLNFGQTGDSSEVP-----DPQPLGEPPA--TPAA 60  
DB 120 KPTKRSKPPPHIFINLAK-----KKKAGAGQVQRDLNAPMSDGAQVQDGGQPAVRNERA 173  
QY 61 VGPPTWASGGGAPMADNNEGADGVGNASGNWH-----CDSTWLGDRVITTTSTWALP 113  
DB 174 TSGNGSGGGG-----GGSGGGVGISTGTFTNNQTEFKLENGWV--EITANSRLVHLN 225  
QY 114 TYNHLYKQI-----SSASTGASNDNHFGYSTPMGYFEENRPHCHFSPRDWORLINN 166  
DB 226 MPESYRVRVNNMDKTAANGNMLADDDIHAQIVTPMSLVANAGWVFNPGDWQLIVNT 285  
QY 167 NMGRFRKRLNFKLFIQVKEVT---TNDGVTTIANNLSTTVQVFSDSYQLPVLGSAHQ 223

```
Db      286 MSSELHVSPEQIFNVVLKTVSESATQPTKVYNDLTASLVALDSNNTMBFTPAAMRS 345
Qy      224 GCL-----PPFPA-----DVFMIPOYGYLTLLNGSQAVGRS-----SFYCL 259
Db      346 ETLGFYFWKPTTPTWRYFYQMDRTLIPSH-----TGTSGPTNIYHGTDPPDDVOQFYTI 399
Qy      260 E-YPPSOMLRTGNF--TFSYTFEEVPFHSSYAHQSQSLDRLMN--PLIDQYLYLNRTONQ 315
Db      400 ENSVPVHLLRTGDEFATGTFDFCKP--CRLTHWTQTNPALGLPP-----FLNSLPQS 450
Qy      316 SGSANQKDLLFGRGSPAGMSVQPKWLPGPCYRQQRVSKTKTDNNNSNF-----TW 366
Db      451 EGATNFGDI-----GV-----QDDKRRGVQMGNTNYITEATIMRPAE 488
Qy      367 TGASKYNLNGRESIINP-----GTAMASHKDDKDEFFPMSGVMIKESAGASNTALD-- 419
Db      489 VGSAPYYSFEASTOGPFKTPIAAGRGGAQTDENQAADGNPRYAFGRHQGKTTTGTETP 548
Qy      420 -----NVMITDEEIKATNPVATERFGTVAVNFOSSSDP 454
Db      549 ERFTYIAHQDTGRYPEGDWQININFLPVTNDNVLLPTDPIG-----GKTGINY--TNIFN 602
Qy      455 ATGDVHANGALPGMWQDRDVLQGPWAKIPIHTDGHFHPSPLMGGFGLK----- 504
Db      603 TYGPLTALNNVP-----PVYPNGQIWDKEFDTD-----LKPRLHVNAFFV 642
Qy      505 ---NPPQILIKNTVPVPAN---PPAEFSATKFASFITQYSTQVSVSEIWELOKENSKEW 558
Db      643 CQNNCEGQLFVKVAPNLTHEYDPDASANMSR-----IVTYSDFWKGKLVFKAKLRASHTW 698
Qy      559 NPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIG 591
Db      699 NPIQQMSIN-----VDNQFNIV-PSNIG 720
```

Search completed: January 21, 2004, 16:02:12  
Job time : 12.2558 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:00:06 ; Search time 31.0877 Seconds  
(without alignments)  
4972.168 Million cell updates/sec

Title: US-09-807-802a-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQPDSSS.....NNGLYTEPRPCTRYLTRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTRMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3251	100.0	736	12 Q9WBP8	Q9wbp8 adeno-associ
2	3229	99.3	736	12 O56137	O56137 adeno-associ
3	2832	87.1	736	12 O56139	O56139 adeno-associ
4	2815	86.6	736	12 Q65311	Q65311 adeno-associ
5	2764.5	85.0	735	12 O56652	O56652 adeno-associ
6	2759.5	84.9	735	12 O56653	O56653 adeno-associ
7	2709.5	83.3	737	12 Q8JQG0	Q8jgg0 adeno-associ
8	2679	82.4	738	12 Q8JQF8	Q8jgf8 adeno-associ
9	2481.5	76.3	533	12 Q92917	Q92917 adeno-associ
10	1830.5	56.3	734	12 O41855	O41855 adeno-associ
11	1717	52.8	732	12 Q67666	Q67666 goose parvo
12	1714	52.7	732	12 Q83290	Q83290 muscovy duc
13	1712	52.7	587	12 Q67667	Q67667 goose parvo
14	1712	52.7	732	12 Q8V395	Q8v395 goose parvo
15	1702	52.4	732	12 Q65444	Q65444 barbarie du
16	1701	52.3	732	12 Q83289	Q83289 muscovy duc

17	1700.5	52.3	724	12 Q9YIJ1	Q9yij1 adeno-associ
18	1698	52.2	534	12 Q67668	Q67668 goose parvo
19	1695	52.1	587	12 Q65445	Q65445 barbarie du
20	1676	51.6	534	12 Q65446	Q65446 barbarie du
21	1540	47.4	676	12 Q67672	Q67672 goose parvo
22	715	22.0	179	12 Q9W8U7	Q9w8u7 duck parvov
23	715	22.0	179	12 Q9WA24	Q9wa24 goose parvo
24	711	21.9	179	12 Q9WN18	Q9wn18 duck parvov
25	709	21.8	179	12 Q9WN19	Q9wn19 duck parvov
26	707	21.7	179	12 Q9WN20	Q9wn20 goose parvo
27	632.5	19.5	703	12 Q8QQV5	Q8qqv5 minute viru
28	618	19.0	571	12 Q8QQV4	Q8qqv4 minute viru
29	545.5	16.8	947	12 Q918U9	Q918u9 bovine parv
30	508	15.6	781	12 Q9PZT0	Q9pzt0 human parvo
31	503.5	15.5	781	12 Q8JNS6	Q8jns6 human parvo
32	501	15.4	554	12 Q9PZ59	Q9pzs9 human parvo
33	501	15.4	781	12 Q9JGP8	Q9jgp8 human parvo
34	500.5	15.4	773	12 Q913X1	Q913x1 human parvo
35	500.5	15.4	781	12 Q8JYE3	Q8jye3 erythroviru
36	499.5	15.4	781	12 Q8JYD9	Q8jyd9 erythroviru
37	499.5	15.4	785	12 Q9J0X4	Q9j0x4 pig-tailed
38	499	15.3	781	12 P89318	P89318 human parvo
39	499	15.3	781	12 P89319	P89319 human parvo
40	498.5	15.3	781	12 P89317	P89317 human parvo
41	497.5	15.3	781	12 Q912B8	Q912b8 human eryth
42	497	15.3	769	12 Q9PZT4	Q9pzt4 human parvo
43	497	15.3	781	12 P90223	P90223 human parvo
44	497	15.3	781	12 Q85191	Q85191 human parvo
45	497	15.3	781	12 P90221	P90221 human parvo

ALIGNMENTS

RESULT 1

Q9WBP8 ID Q9WBP8 PRELIMINARY; PRT; 736 AA.  
AC Q9WBP8;  
CT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE Capsid protein.  
OS Adeno-associated virus 1.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=85106;  
RN [1]\_TaxID=85106;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9214338; PubMed=10196295;  
RA Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;  
RT "Gene therapy vectors based on adeno-associated virus type 1.";  
RL J. Virol. 73:3994-4003(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Xiao W., Wilson J.M.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF063497; AAD27757.1; -;  
DR InterPro; IPR001403; Parvo.coat.  
DR Pfam; PF00740; Parvo.coat; -;  
SQ SEQUENCE 736 AA; 81375 MW; CFABP89B5CD0595 CRC64;

Query Match 100.0%; Score 3251; DB 12; Length 736;  
Best Local Similarity 100.0%; Pred. No. 3.9e-227;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TAPGKKRPVEQSPQPDSSSGIGTKTQGPAPKRLNFGTGDSVDDPQPLGEPATPAA	60
Db	138	TAPGKKRPVEQSPQPDSSSGIGTKTQGPAPKRLNFGTGDSVDDPQPLGEPATPAA	197
Qy	61	VGPTTASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY	120
Db	198	VGPTTASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY	257
Qy	121	KOISSASTGASNDNHYFGYSTPWGYDFNRFCHFSRDMQRLNINNNNGFRPKRLNFKLF	180

```

Db 258 KQISSASTGASNDNHFGYSTPMGYFDNFRPHCHFSPRDWQRLNNWNGFRPKRLNFKLF 317
QY 181 NIQVKEVTNDGVTTIANNLTSVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOYG 240
Db 318 NIQVKEVTNDGVTTIANNLTSVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOYG 377
QY 241 YLTLLNGSAQGRSSFYCLEYFPPSOMLRTGNNTFTSYTPEEVPFHSSVAHSQSLODLRMLNP 300
Db 378 YLTLLNGSAQGRSSFYCLEYFPPSOMLRTGNNTFTSYTPEEVPFHSSVAHSQSLODLRMLNP 437
QY 301 LIDQYLYLNRNTQNGSQAQNKDILLFSRGSAGMSVQPKWLPQPCYRQORVSKTKTDNN 360
Db 438 LIDQYLYLNRNTQNGSQAQNKDILLFSRGSAGMSVQPKWLPQPCYRQORVSKTKTDNN 497
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDN 420
Db 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDN 557
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480
Db 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 617
QY 481 IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTG 540
Db 618 IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTG 677
QY 541 QVSVEIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNGLYTEPRPIGRYLTRPL 599
Db 678 QVSVEIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNGLYTEPRPIGRYLTRPL 736

```

## RESULT 2

```

O56137 PRELIMINARY; PRT; 736 AA.
ID O56137
AC O56137
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DE Capsid protein VPI.
OS Adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
(AAV) serotypes other than AAV type 2.";
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF028704; AAB95450.1; -;
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;

```

```

Query Match 99.3%; Score 3229; DB 12; Length 736;
Best Local Similarity 99.2%; Pred. No. 1.5e-225;
Matches 594; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TAPGKRPRVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPQPLGEPATPAA 60
Db 138 TAPGKRPRVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPQPLGEPATPAA 197
QY 61 VGPTTASGGAPMADNNEGADGVGNASGNHCDSTWLGDRTTSTRTWALPTVNNHLY 120
Db 198 VGPTTASGGAPMADNNEGADGVGNASGNHCDSTWLGDRTTSTRTWALPTVNNHLY 257
QY 121 KQISSASTGASNDNHFGYSTPMGYFDNFRPHCHFSPRDWQRLNNWNGFRPKRLNFKLF 180

```

```

Db 258 KQISSASTGASNDNHFGYSTPMGYFDNFRPHCHFSPRDWQRLNNWNGFRPKRLNFKLF 317
QY 181 NIQVKEVTNDGVTTIANNLTSVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOYG 240
Db 318 NIQVKEVTNDGVTTIANNLTSVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPOYG 377
QY 241 YLTLLNGSAQGRSSFYCLEYFPPSOMLRTGNNTFTSYTPEEVPFHSSVAHSQSLODLRMLNP 300
Db 378 YLTLLNGSAQGRSSFYCLEYFPPSOMLRTGNNTFTSYTPEEVPFHSSVAHSQSLODLRMLNP 437
QY 301 LIDQYLYLNRNTQNGSQAQNKDILLFSRGSAGMSVQPKWLPQPCYRQORVSKTKTDNN 360
Db 438 LIDQYLYLNRNTQNGSQAQNKDILLFSRGSAGMSVQPKWLPQPCYRQORVSKTKTDNN 497
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDN 420
Db 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDN 557
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480
Db 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 617
QY 481 IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTG 540
Db 618 IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTG 677
QY 541 QVSVEIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNGLYTEPRPIGRYLTRPL 599
Db 678 QVSVEIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNGLYTEPRPIGRYLTRPL 736

```

## RESULT 3

```

O56139 PRELIMINARY; PRT; 736 AA.
ID O56139
AC O56139
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DE Capsid protein VPI.
OS Adeno-associated virus 3B.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68742;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
(AAV) serotypes other than AAV type 2.";
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF028705; AAB95452.1; -;
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81906 MW; DD52331AD5F0D70F CRC64;

```

```

Query Match 87.1%; Score 2832; DB 12; Length 736;
Best Local Similarity 85.7%; Pred. No. 9.1e-197;
Matches 514; Conservative 34; Mismatches 50; Indels 2; Gaps 2;

```

```

QY 1 TAPGKRPRVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPQPLGEPATPAA 60
Db 138 TAPGKRPRVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPQPLGEPATPAA 197
QY 61 VGPTTASGGAPMADNNEGADGVGNASGNHCDSTWLGDRTTSTRTWALPTVNNHLY 120
Db 198 LGSNTMASGGAPMADNNEGADGVGNSSGNHCDSTWLGDRTTSTRTWALPTVNNHLY 257
QY 121 KQISSASTGASNDNHFGYSTPMGYFDNFRPHCHFSPRDWQRLNNWNGFRPKRLNFKLF 180
Db 258 KQISSASTGASNDNHFGYSTPMGYFDNFRPHCHFSPRDWQRLNNWNGFRPKRLNFKLF 316

```



```

QY 181 NIQKVTNDGTTIANNLTSTVQVSDSEYQLPYVLGSAHQGLPPFPADVFMIPQYG 240
DB 317 NIQKVTNDGTTIANNLTSTVQVSDSEYQLPYVLGSAHQGLPPFPADVFMIPQYG 376
QY 241 YLTNNGSAQGRSSFYCYLEYPPSOMLRGNFTPSYTFEEVPHSSYAHQSOLDRLMNP 300
DB 377 YLTNNGSAQGRSSFYCYLEYPPSOMLRGNFTPSYTFEEVPHSSYAHQSOLDRLMNP 436
QY 301 LIDQYLYLNRTQ-NQSGSAQNKDLFLSRGSPAGMSVQPKWLPGPCYRQQRVSKTKTDN 359
DB 437 LIDQYLYLNRTQ-NQSGSAQNKDLFLSRGSPAGMSVQPKWLPGPCYRQQRVSKTKTDN 496
QY 360 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKFFPMGCVMIFGKESAGASNTALD 419
DB 497 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKFFPMGCVMIFGKESAGASNTALD 556
QY 420 NMVITDEBEIKATNPVATERFGTVANFQSSSDPATGVDVHANGALPGMWQDRDYLQ 479
DB 557 NMVITDEBEIKATNPVATERFGTVANFQSSSDPATGVDVHANGALPGMWQDRDYLQ 616
QY 480 PIWAKIPHTDGHFHPSPLMGGFGLKXNPPQILIKNTVPANPPAEPKATKFASTOYST 539
DB 617 PIWAKIPHTDGHFHPSPLMGGFGLKXNPPQILIKNTVPANPPAEPKATKFASTOYST 676
QY 540 GOVSVEIEWELQKENSKRWNPEVQVTSYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599
DB 677 GOVSVEIEWELQKENSKRWNPEVQVTSYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 4
Q65311 PRELIMINARY; PRT; 736 AA.
ID AC Q65311;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE Capsid protein.
OS Adeno-associated virus 3.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=46350;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RX MEDLINE=96266430; PubMed=8661429;
RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
RT "Nucleotide sequencing and generation of an infectious clone of adeno-
associated virus 3.";
RL Virology 221:208-217(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RA Muramatsu S., Brown K.E.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U48704; AAC5049.1; -.
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81660 MW; AFFIEF47B5C67A10 CRC64;

Query Match 86.6%; Score 2815; DB 12; Length 736;
Best Local Similarity 85.2%; Pred. No. 1.6e-195;
Matches 511; Conservative 35; Mismatches 52; Indels 2; Gaps 2;

QY 1 TAPGKKRPVEQSPQEDSSSGIGKTCQPAKRLNFGQTDGSEVDPDQPLGEPATPAA 60
DB 138 TAPGKKGAVDQSPQEDSSSGVGSCKQPAKRLNFGQTDGSEVDPDQPLGEPAPPTS 197

QY 61 VQPTTASGGGAPMADNNGAGVGNAGNHCDSFWLGDRTVITSTRTWALPTYNHLY 120
DB 198 LGSNTWASGGGAPMADNNGAGVGNAGNHCDSFWLGDRTVITSTRTWALPTYNHLY 257

QY 121 KOISSASTGASNDNHFGYSTPKGYDFDNRHCHFSPRDQRLNNWGFPRKRLNFKLF 180

```

```

DB 258 KQ18SQS-GASNDNHFGYSTPKGYDFDNRHCHFSPRDQRLNNWGFPRKRLNFKLF 316
QY 181 NIQKVTNDGTTIANNLTSTVQVSDSEYQLPYVLGSAHQGLPPFPADVFMIPQYG 240
DB 317 NIQKVTNDGTTIANNLTSTVQVSDSEYQLPYVLGSAHQGLPPFPADVFMIPQYG 376
QY 241 YLTNNGSAQGRSSFYCYLEYPPSOMLRGNFTPSYTFEEVPHSSYAHQSOLDRLMNP 300
DB 377 YLTNNGSAQGRSSFYCYLEYPPSOMLRGNFTPSYTFEEVPHSSYAHQSOLDRLMNP 436
QY 301 LIDQYLYLNRTQ-NQSGSAQNKDLFLSRGSPAGMSVQPKWLPGPCYRQQRVSKTKTDN 359
DB 437 LIDQYLYLNRTQ-NQSGSAQNKDLFLSRGSPAGMSVQPKWLPGPCYRQQRVSKTKTDN 496
QY 360 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKFFPMGCVMIFGKESAGASNTALD 419
DB 497 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKFFPMGCVMIFGKESAGASNTALD 556
QY 420 NMVITDEBEIKATNPVATERFGTVANFQSSSDPATGVDVHANGALPGMWQDRDYLQ 479
DB 557 NMVITDEBEIKATNPVATERFGTVANFQSSSDPATGVDVHANGALPGMWQDRDYLQ 616
QY 480 PIWAKIPHTDGHFHPSPLMGGFGLKXNPPQILIKNTVPANPPAEPKATKFASTOYST 539
DB 617 PIWAKIPHTDGHFHPSPLMGGFGLKXNPPQILIKNTVPANPPAEPKATKFASTOYST 676
QY 540 GOVSVEIEWELQKENSKRWNPEVQVTSYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599
DB 677 GOVSVEIEWELQKENSKRWNPEVQVTSYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 5
O56652 PRELIMINARY; PRT; 735 AA.
ID AC O56652;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Major coat protein VP1.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
proteins affect viral infectivity: lack of an RGD integrin-binding
motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF043303; AAC03780.1; -.
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
KW Coat protein.
FT VARIANT 76 76 D -> V.
FT VARIANT 553 553 D -> N.
FT VARIANT 567 567 T -> N.
FT VARIANT 677 678 QV -> HV.
FT VARIANT 710 710 V -> R.
SQ SEQUENCE 735 AA; 81944 MW; 980BEEF46908390B CRC64;

Query Match 85.0%; Score 2764.5; DB 12; Length 735;
Best Local Similarity 82.8%; Pred. No. 7.1e-192;
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKKRPVEQSPQEDSSSGIGKTCQPAKRLNFGQTDGSEVDPDQPLGEPATPAA 60

```



```

Qy 1  TAPGKRPVPSQPO - EPDSSSGIGKTCQPAKRLNFGQDGSSEVPDPQPLGEPATPA 59
Db 138 TAPAKRPVPSQPS - PDSSSTGIGKGCQPAKRLNFGQDGSSEVPDPQPLGEPAPAS 197
Qy 60  AVGPTTASGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTRTWALPTVNNHL 119
Db 198 SVSGTVAAGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTRTWALPTVNNHL 257
Qy 120 YKOISSASTGASNDNHFFGYSTPMGYEDNRFCHFSFPRDWORLNNWGFPRKRLNFKL 179
Db 258 YKOISSASTGASNDNHFFGYSTPMGYEDNRFCHFSFPRDWORLNNWGFPRKRLNFKL 317
Qy 180 FNIQKVEVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQY 239
Db 318 FNIQKVEVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQY 377
Qy 240 GYLTLNNGSOAVGRSSFYCLEYFPFQMLRTGNFTFSTFEEVPHSSYAHQSQSLDRLM 299
Db 378 GYLTLNNGSOAVGRSSFYCLEYFPFQMLRTGNFTFSTFEEVPHSSYAHQSQSLDRLM 437
Qy 300 PLIDQYLYLNRTQ - NOSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTD 358
Db 438 PLIDQYLYLNRTQSNPGGTAGNRELQFYGGSTMAEQAKNWLPGPCYRQQRVSKTKTD 497
Qy 359 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIIPKESAGASN - TA 417
Db 498 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIIPKESAGASN - TA 555
Qy 418 LDNWMITDEBEIKATNPVATERGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYL 477
Db 556 LDNWMITDEBEIKATNPVATERGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYL 615
Qy 478 QGIWAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTVPANPPAEFSATKPSFTQY 537
Db 616 QGIWAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTVPANPPAEVFTPAKASFTQY 675
Qy 538 STGQSVSEIWELOKENSKRNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTR 597
Db 676 STGQSVSEIWELOKENSKRNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTR 735
Qy 598 PL 599
Db 736 NL 737

RESULT 8
Q8JQF8 PRELIMINARY; PRT; 738 AA.
AC Q8JQF8
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Capsid protein.
OS Adeno-associated virus 8.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=202813;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12192090;
RA Gao G.P., Alvira M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
RL human gene therapy."
RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859(2002).
DR EMBL; AF513852; AAN03857.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat_1.
SQ SEQUENCE 738 AA; 81756 MW; 59BC2BE1B148AA14 CRC64;

Query Match 82.4%; Score 2679; DB 12; Length 738;
Best Local Similarity 80.7%; Pred. No. 1.1e-185;
Matches 485; Conservative 47; Mismatches 67; Indels 2; Gaps 2;

Qy 1  TAPGKRPVPSQPO - EPDSSSGIGKTCQPAKRLNFGQDGSSEVPDPQPLGEPATPA 59

```

```

Db 138 TAPGKRPVPSQPS - PDSSSTGIGKGCQPAKRLNFGQDGSSEVPDPQPLGEPAPAS 197
Qy 60  AVGPTTASGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTRTWALPTVNNHL 119
Db 198 SVSGTVAAGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTSTRTWALPTVNNHL 257
Qy 120 YKOISSASTGASNDNHFFGYSTPMGYEDNRFCHFSFPRDWORLNNWGFPRKRLNPK 178
Db 258 YKOISSASTGASNDNHFFGYSTPMGYEDNRFCHFSFPRDWORLNNWGFPRKRLNPK 317
Qy 179 FNIQKVEVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQ 238
Db 318 FNIQKVEVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQ 377
Qy 239 GYLTLNNGSOAVGRSSFYCLEYFPFQMLRTGNFTFSTFEEVPHSSYAHQSQSLDRLM 298
Db 378 GYLTLNNGSOAVGRSSFYCLEYFPFQMLRTGNFTFSTFEEVPHSSYAHQSQSLDRLM 437
Qy 299 NPLIDQYLYLNRTQ - NOSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTD 358
Db 438 NPLIDQYLYLNRTQSNPGGTAGNRELQFYGGSTMAEQAKNWLPGPCYRQQRVSKTKTD 497
Qy 359 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIIPKESAGASN - TA 418
Db 498 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIIPKESAGASN - TA 557
Qy 419 LDNWMITDEBEIKATNPVATERGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYL 478
Db 558 LDNWMITDEBEIKATNPVATERGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYL 617
Qy 479 GGIWAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTVPANPPAEFSATKPSFTQY 538
Db 618 GGIWAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTVPANPPAEVFTPAKASFTQY 677
Qy 539 STGQSVSEIWELOKENSKRNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTR 598
Db 678 STGQSVSEIWELOKENSKRNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTR 737
Qy 599 L 599
Db 738 L 738

RESULT 9
O92917 PRELIMINARY; PRT; 533 AA.
ID O92917
AC O92917
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major coat protein VP3.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9508582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RL proteins affect viral infectivity: lack of an RGD integrin-binding
RL motif."
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043303; AAC03779.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat_1.
KW Coat protein.

```

FT VARIANT 351 351 D -> N.  
 FT VARIANT 365 365 T -> N.  
 FT VARIANT 475 476 QV -> HV.  
 FT VARIANT 508 508 V -> R.  
 SQ SEQUENCE 533 AA; 60063 MW; 9E4D8BC25810D4F0 CRC64;  
 Query Match 76.3%; Score 2481.5; DB 12; Length 533;  
 Best Local Similarity 83.3%; Pred. No. 1.5e-171;  
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
 QY 66 MASGGAPMADNNEGADGVGNASGNHWCDSWLGDRVITTSRTWALPTYNHLYKQISS 125  
 DB 1 MATSGAPMADNNEGADGVGNSSGNHWCDSWLGDRVITTSRTWALPTYNHLYKQISS 60  
 QY 126 ASTGASNDNHFGYSTPWCYFDFNRFCHFSFRDQRLNNNNGFRPKRLNFKLNIQVK 185  
 DB 61 QS-GASNDNHFGYSTPWCYFDFNRFCHFSFRDQRLNNNNGFRPKRLNFKLNIQVK 119  
 QY 186 EVTTNDGVTIANNLTSVQVPSDSYQLPYVLGSAHQGCLPPFPADVPFMIQYGYLTJN 245  
 DB 120 EVTQNDGTTIANNLTSVQVPSDSYQLPYVLGSAHQGCLPPFPADVPFMIQYGYLTJN 179  
 QY 246 NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEVPFPHSSVYAHQSQSLDR 305  
 DB 180 NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEVPFPHSSVYAHQSQSLDR 239  
 QY 306 LYLINRTQSGSAQNKDLFRSGSPAGMSVQPKWLPGPCYRQQRVSKTKTDNNNSNPT 365  
 DB 240 LYLISRTNPSTGTTQSRLOFSQAGASDIRDQSRNWLPGCYRQQRVSKTSADNNSEYS 299  
 QY 366 WTGASKYNLNGRESINPGTAMASHKDEKDFPFMSGVMI PKESAGASANTALDNVMTD 425  
 DB 300 WTGATKYHLNGRSLVNPQPMASHKDEKDFPFQSGVLIFKQSGSEKTNVDIEKVMITD 359  
 QY 426 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHMGALPGMWQDRDVLQGPWAKI 485  
 DB 360 EEEIRTNPVATEQYGSVSNLQGRNQRAADVNTQGVLPQMWQDRDVLQGPWAKI 419  
 QY 486 PHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPAFSAKFASFITQYSTGVSV 545  
 DB 420 PHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPSTTFAAKFASFITQYSTGVSV 479  
 QY 546 IEWELOKENSKRNPEVQYTSYAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 599  
 DB 480 IEWELOKENSKRNPEIQYTSYNTKSNVNDFTVDTNGVYSEPRPIGTRYLTRPL 533  
 RESULT 10  
 O41855 PRELIMINARY; PRT; 734 AA.  
 AC O41855  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Capsid.  
 OS Adeno-associated virus 4.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
 OX NCBI\_TaxID=57579;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC VR-646;  
 RX MEDLINE=97404695; PubMed=9261407;  
 RA Chiorini J.A., Yang L., Liu Y., Safer B., Kotin R.M.;  
 RT "Cloning of adeno-associated virus type 4 (AAV4) and generation of  
 recombinant AAV4 particles";  
 RL J. Virol. 71:6823-6833(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC VR-646;  
 RA Chiorini J.A., Yang L., Kotin R.M., Safer B.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 DR ENBL; U89790; AAC58045.1; --  
 DR InterPro; IPR001403; Parvo\_coat.

DR Pfam; PF00740; Parvo\_coat; 1.  
 SQ SEQUENCE 734 AA; 80639 MW; 616CC27A77BB6F CRC64;  
 Query Match 56.3%; Score 1830.5; DB 12; Length 734;  
 Best Local Similarity 57.8%; Pred. No. 3.6e-124; Indels 31; Gaps 10;  
 Matches 355; Conservative 74; Mismatches 154;  
 QY 1 TAPGKRPVEQSPQEPDSSGIGTKGQPAKRLNF-GQTGDSVSPDPQPLGPPATPA 59  
 DB 137 TAPGKRPLIESPQPDSSSTGIGKKGPAKRLVFEDETGAAGDPPEGSTG-----A 190  
 QY 60 AVGTTWASGGAPMADNNEGADGVGNASGNHWCDSWLGDRVITTSRTWALPTYNHLY 119  
 DB 191 MSDISEKRAAAGAAVGGQAGGVGNASGDWHDCTSWSEGHVTTTSTRTWALPTYNHLY 250  
 QY 120 YKOISASTGASNDNHFGYSTPWCYFDFNRFCHFSFRDQRLNNNNGFRPKRLNFKL 179  
 DB 251 YKRLGE-----SQSNYNGFSTPWCYFDFNRFCHFSFRDQRLNNNNGFRPKRAVKI 306  
 QY 180 FNIQVKEVTTNDGVTIANNLTSVQVPSDSYQLPYVLGSAHQGCLPPFPADVPFMIQY 239  
 DB 307 FNIQVKEVTTNGETTVANNLTSTVQIFADSSYELFYVMDAGQSGSLPPFPFNDVFMVQY 366  
 QY 240 GY---LTLNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEVPFPHSSVYAHQSQSLDR 296  
 DB 367 GYCGLVGTNTSQOQTDNRNFAFYCLEYFSPQMLRTGNFTFSYTFEVPFPHSSVYAHQSQSLDR 426  
 QY 297 LWNPLIDQLYLYLNRQNTQ-----QSGSAQNKDLFRSGSPAGMSVQPKWLPGPCYRQQRV 352  
 DB 427 LWNPLIDQLYLNGLOSTTGTTLNAGTATTN---ETKLRTNFSNPKWLPGPCYRQQRV 483  
 QY 353 SKTKTDNNNSNFTWTGAS---KY---NLNGRESINPGTAMASHKDEKDFPFMSGVMI 405  
 DB 484 SKTA--NQYKIPATGSDSLIKYETHSTLDGRWSALTGPPMATAGPADSK-FSNSQLIF 540  
 QY 406 PKESAGASANTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHMGAL 465  
 DB 541 APGKONGTATVPGLIFTSEELAATNATDTDMWNLPGGQDSNLSLPTVDRLTALGAV 600  
 QY 466 PGMWQDRDVLQGPWAKI PHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPAF 525  
 DB 601 PGMWQDRDIYQGPWAKI PHTDGHFHPSPMLGGFGLKHPPPQIFIKNTVPANPAT 660  
 QY 526 SATKFPASFITQYSTGVSVIEWELOKENSKRNPEVQYTSYAKSANVDFTVNNGLYT 585  
 DB 661 SSTPVNSFITQYSTGVSVQVQIDWEIQKRSKRNPEVQYTSYAKSANVDFTVNNGLYT 720  
 QY 586 EPRPIGTRYLTRPL 599  
 DB 721 EPRPIGTRYLTRPL 734  
 RESULT 11  
 O67666 PRELIMINARY; PRT; 732 AA.  
 AC O67666  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE VP1.  
 GN VP1.  
 OS Goose parvovirus.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=38251;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Virus B;  
 RX MEDLINE=96010229; PubMed=7571426;  
 RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
 RT "Analysis of the complete nucleotide sequences of goose and muscovy  
 duck parvoviruses indicates common ancestral origin with adeno-  
 associated virus 2";  
 RL Virology 212:562-573 (1995).

```

RX MEDLINE=96406928; PubMed=8811015;
RA Le Gall-Recule G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;
RT "Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)
in a baculovirus expression system and demonstration of immunity
induced by the recombinant protein.";
RL J. Gen. Virol. 77:2159-2163(1996).
DR EMBL; U25749; AAA83230.1; -.
DR InterPro: IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat; 1.
DR EMBL; 268272; CAA92575.1; -.
DR InterPro: IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat; 1.
FT CHAIN 146 732
FT CHAIN 199 732
FT CHAIN 199 732
SQ SEQUENCE 732 AA; 81364 MW; DE70DCAB215F4E2 CRC64;

Query Match 52.8%; Score 1714; DB 12; Length 732;
Best Local Similarity 53.0%; Pred. No. 6.1e-116;
Matches 327; Conservative 94; Mismatches 148; Indels 48; Gaps 13;

QY 1 TAPGKKR-----VPEQSPQ-EPDSSSGIGKGTGOAPAKRLNFQGTGDSVPDPQL 51
DB 146 TAPAKKNTGKLTTHYFVVKPKLUTEEVSAGGSAVQDG-----GATAE----- 189
QY 52 GEPPTAAVGPPTMASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTWA 111
DB 190 GTEP-----VAASEMAEGGGAGDSSGGADGVGNAGNWHCDSQMGNVTVIKTRTW 244
QY 112 LPTNNHLYKQISSASTGASNDNH--YFGYSTPWGFDNRFCHFSPRDWQRLNNWG 169
DB 245 LPSYNNHIYKAITS---GTSQDANVQAGYSTPWGYDFENRFCHFSPRDWQRLNNWG 301
QY 170 FRPKLNFKLFNIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPVYLGSAHQGLPPF 229
DB 302 IRPKSLKPLFNVQKEVTTDQTKTIANNLTSTIQTVDDEHQLPVYLGSAHQGLPPF 361
QY 230 PADVFMIPQGYLTN---NGSOAVGRSSPYCLEYFPSSQMLRTGNNFTSYTFEVPFHS 286
DB 362 PSDVYALPOGYCTMHTNQGARNFDRSAFYCLEYFPSSQMLRTGNNFETFEVPFHS 421
QY 287 SYAHSQSLRLMPLDQYLYLNRNTQNGSGAQNKDLFLSRGSPAGMSVQPKWLPGPC 346
DB 422 MFAHSQDLRLMPLDQYLYLNFNEV--DSSRNAQ-----FKKAVGAYGTMGRENWLP 475
QY 347 YRQRV-SKTKTNNNSNFT-WTGASKYNLNGRESINPGTAMASHKDDKFPMSGYM 404
DB 476 FLQVRVAYTGGTDNANWINGKNVNLKQRYLLQPGVSATYTEGEASLPAQNIL 535
QY 405 IFQKE--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFOSSTDPATGDVHAM 462
DB 536 GIAKDPVRSSTTAGISDIWVTEEQEVAFTNGVGVKPYGRVTVNEQNTTTAPTSSDL 595
QY 463 GALPGMWQDRDYVLOGPIWAKIPHTDGHFHPSPMLGGFGLKNPPQIILKNTVPANPP 522
DB 596 GALPGMWQNRDIYLOGPIWAKIPKTDGKFHPSNLGGFGLHNPPQVFIKNTVPADPP 655
QY 523 APTSATKFAFITQYSTGQVSVEIWELOKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 582
DB 656 VEYVHQWNSYITQYSTGQCTVEMWELKRENSKRWNPEIQFTSNFNRSTINFAFNETG 715
QY 583 LYTEPRIPGRYLTPL 599
DB 716 GYVEDRLIGRYLTQNL 732

RESULT 12
Q83290
ID Q83290 PRELIMINARY; PRT; 732 AA.
AC Q83290;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Capsid protein.
OS Muscovy duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=37325;
RN [1]
RP STRAIN=89384/France;

RX MEDLINE=96406928; PubMed=8811015;
RA Le Gall-Recule G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;
RT "Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)
in a baculovirus expression system and demonstration of immunity
induced by the recombinant protein.";
RL J. Gen. Virol. 77:2159-2163(1996).
DR EMBL; 268272; CAA92575.1; -.
DR InterPro: IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat; 1.
DR EMBL; 268272; CAA92575.1; -.
DR InterPro: IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat; 1.
FT CHAIN 146 732
FT CHAIN 199 732
FT CHAIN 199 732
SQ SEQUENCE 732 AA; 81364 MW; DE70DCAB215F4E2 CRC64;

Query Match 52.7%; Score 1714; DB 12; Length 732;
Best Local Similarity 53.3%; Pred. No. 1e-115;
Matches 327; Conservative 82; Mismatches 164; Indels 40; Gaps 11;

QY 6 KRPVQSPQEPDSSSGIGK-TGQPAKRLNFQGTGDSVPDPQLG-----E 53
DB 141 BEPVNTAPAKKSS---GKLTTHYFIVKPKLSE---ENSPSPNSGGEASAAATEGSE 192
QY 54 PPATPAAVGPPTMASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTWA 113
DB 193 PVAAP-----NMAEGSGGAGDSSGGADGVGNAGNWHCDSQMGNVTVIKTRTW 246
QY 114 TYNHLYKQISSASTGASNDNHYFGYSTPWGFDNRFCHFSPRDWQRLNNWGFRPK 173
DB 247 SYNNHIYKAITSGTWPSN-TQYAGYSTPWGFDNRFCHFSPRDWQRLNNWGIRPK 305
QY 174 RLNFKLFNIQVKEVTNDGVTTIANNLTSTVQFSDSEYQLPVYLGSAHQGLPPPADV 233
DB 306 ALKFKIFNVQKEVTTDQTKTIANNLTSTIQTVDNEHQLPVYLGSAHQGLPPPADV 365
QY 234 FMIPQGYLTN---NGSOAVGRSSPYCLEYFPSSQMLRTGNNFTSYTFEVPFHSYAH 290
DB 366 YALPOGYCTMHTNQGARNFDRSAFYCLEYFPSSQMLRTGNNFETFEVPFHSMPAH 425
QY 291 QSQSLRLMPLDQYLYLNRNTQNGSGAQNKDLFLSRGSPAGMSVQPKWLPGPCYRQ 350
DB 426 SQDLRLMPLDQYLYLNFSEV-NGRNAQ-----PKKAVGAFGAMGNWLPGLKLDQ 479
QY 351 RVSKTK--TDNNNSNFTWTGASKYNLNGRESINPGTAMASHKDDKFPMSGYMIFOK 408
DB 480 RVRAVSGGTDNANWINGKNVNLKQRYLLQPGVSATYTEGEASLPAQNIIIAK 539
QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFOSSTDPATGDVHAMGALP 466
DB 540 DYPYRSGTLAGISDIWVTEEQEVAFTNGVGVKPYGRVTVNEQNTTTAPTNALEVL 599
QY 467 GMVWQDRDYVLOGPIWAKIPHTDGHFHPSPMLGGFGLKNPPQIILKNTVPANPPAEFS 526
DB 600 GMVWQNRDIYLOGPIWAKIPKTDGKFHPSNLGGFGLHNPPQVFIKNTVPADPPLEV 659
QY 527 ATKPASFITQYSTGQVSVEIWELOKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTE 586
DB 660 NQWNSYITQYSTGQCTVEMWELKRENSKRWNPEIQFTSNFNRSTINFAFNETGYE 719
QY 587 PRPIGRYLTPL 599
DB 720 DRLIGRYLTQNL 732

RESULT 13
Q67667
ID Q67667 PRELIMINARY; PRT; 587 AA.
AC Q67667;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE VP2.
GN VP2.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.

```

OX NCBI\_TaxID=38251;  
RN SEQUENCE FROM N.A.  
RC STRAIN=Virulent B;  
RX MEDLINE=96010229; PubMed=7571426;  
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
RT "Analysis of the complete nucleotide sequences of goose and muscovy  
RT duck parvoviruses indicates common ancestral origin with adeno-  
RL associated virus 2";  
RL Virology 212:562-573 (1995).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Virulent B;  
RA Zadori Z.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25749; AA083231.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 587 AA; 65246 MW; 0278E5D5FE7F0423 CRC64;

Query Match 52.7%; Score 1712; DB 12; Length 587;  
Best Local Similarity 52.9%; Pred. No. 18-115;  
Matches 326; Conservative 94; Mismatches 148; Indels 48; Gaps 13;

QY 2 AFGKRR-----PVEQSPQ-EPDSSSGIGKTGQPAKRLNFGQTGDSSEVPDPQLG 52  
DB 2 APAKNTGKLTDRHPVVKPKLTETEEVSAGGSSAVQDG-----GATAE-----G 45

QY 53 EPPATPAACPTTMASGGGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTSRTWA 112  
DB 46 TEP-----VAASEMAGGGGAGDGGADGVNAGNWHCDSTWLGDRVITTSRTWA 100

QY 113 PTNNHLYKQISSASTGASNDH--YFGYSTPWGYDFNRFCHSPRDQRLNINNWG 170  
DB 101 PSNNHLYKAITs---GTSQDANVQAGYSTPWGYDFNRFCHSPRDQRLNINHWG 157

QY 171 RPRLNFKLPIQVKEVTTNDGVTIANNLTSTVQVPSSEYQLPYVLSAHQGLPPFP 230  
DB 158 RPKSLKFKIFNVQVKEVTTNDGVTIANNLTSTVQVPSSEYQLPYVLSAHQGLPPFP 217

QY 231 ADVFMIPOGYLTN---NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSTYEEVPHSS 287  
DB 218 SDVIALPOGYCTMHTNQGARNDRSAFYCLBYFSPQMLRTGNFTFSTYEEVPHSS 277

QY 288 YAHQSQDLRLNPLIDQYLYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPGPCY 347  
DB 278 FAHSQDLRLNPLIDQYLYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPGPCY 331

QY 348 RQORV-SKTKTDNNNSNFT-WTGASKYNLNGRESINPGTAMASHKDDKDFPMSGVMI 405  
DB 332 LDQVRVAYTGGTDYANMNWISNGNKNVNLKDRQYLLQPGVPSATVTEGEASSLPANILG 391

QY 406 FGKE--SAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAM 463  
DB 392 IAKDPYRSGSTTAGISDIWVDEQVAPTNVGWKPYGKTVTNEQNTTAPTSSDLVLG 451

QY 464 ALPGMWQDRDYVLOGPIWAKIPHTDGHFHPSPLMGGFGLKPNPPQILIKNTVPANPPA 523  
DB 452 ALPGMWQNRDIYLOGPIWAKIPKTDGKPHSPNLGGFGLNPPQVFIKNTVPADPPV 511

QY 524 EFSATKFASTIQTQYSTGVSEVLEWELQKENSKRWNPEVQYTSNYSKANSANVDTVDNGL 583  
DB 512 EYVHKWNSYITQYSTGQCTVEMVWELRKENSKRWNPEIQTSNYSKANSANVDTVDNGL 571

QY 584 YTEPRPIGTRYLTP 599  
DB 572 YVEDRLIGTRYLTQNL 587

RESULT 14

QBV395

ID QBV395

AC QBV395;

PRELIMINARY; PRT; 732 AA.

Q65444

ID Q65444

AC Q65444;

DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Capsid protein VP.  
GN Goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=38251;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=GPV-YG;  
RA Ge Y., You Y., Xu Q.;  
RT "Analysis of the major open reading frames' nucleotide sequences in  
RT Goose parvovirus GPV-YG strain isolated in China";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF416726; AAL37722.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 732 AA; 81456 MW; 73F2B4BC769744B6 CRC64;

Query Match 52.7%; Score 1712; DB 115; Length 732;  
Best Local Similarity 53.2%; Pred. No. 1.4e-115;  
Matches 328; Conservative 89; Mismatches 152; Indels 48; Gaps 13;

QY 1 TAPGKRR-----PVEQSPQ-EPDSSSGIGKTGQPAKRLNFGQTGDSSEVPDPQL 51  
DB 146 TAPVKNTGKLTDRHPVVKPKLTETEEVSAGGTSVAVQDG-----GATAE----- 189

QY 52 EPPATPAACPTTMASGGGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTSRTWA 111  
DB 190 GTEP-----VAASEMAGGGGAGDGGADGVNAGNWHCDSTWLGDRVITTSRTWA 244

QY 112 LPTNNHLYKQISSASTGASNDN--HYFGYSTPWGYDFNRFCHSPRDQRLNINNWG 169  
DB 245 LPSNNHLYKAITs---GTSQDANVQAGYSTPWGYDFNRFCHSPRDQRLNINHWG 301

QY 170 RPRLNFKLPIQVKEVTTNDGVTIANNLTSTVQVPSSEYQLPYVLSAHQGLPPFP 229  
DB 302 IRPKSLKFKIFNVQVKEVTTNDGVTIANNLTSTVQVPSSEYQLPYVLSAHQGLPPFP 361

QY 230 ADVFMIPOGYLTN---NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSTYEEVPHSS 286  
DB 362 SDVIALPOGYCTMHTNQGARNDRSAFYCLBYFSPQMLRTGNFTFSTYEEVPHSS 421

QY 287 YAHQSQDLRLNPLIDQYLYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPGPC 346  
DB 422 FAHSQDLRLNPLIDQYLYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPGPC 475

QY 347 RQORV-SKTKTDNNNSNFT-WTGASKYNLNGRESINPGTAMASHKDDKDFPMSGVMI 404  
DB 476 LDQVRVAYTGGTDYANMNWISNGNKNVNLKDRQYLLQPGVPSATVTEGEASSLPANIL 535

QY 405 FGKE--SAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAM 462  
DB 536 IAKDPYRSGSTTAGISDIWVDEQVAPTNVGWKPYGKTVTNEQNTTAPTSSDLVL 595

QY 463 ALPGMWQDRDYVLOGPIWAKIPHTDGHFHPSPLMGGFGLKPNPPQILIKNTVPANPP 522  
DB 596 ALPGMWQNRDIYLOGPIWAKIPKTDGKPHSPNLGGFGLNPPQVFIKNTVPADPP 655

QY 524 EFSATKFASTIQTQYSTGVSEVLEWELQKENSKRWNPEVQYTSNYSKANSANVDTVDNNG 582  
DB 656 EYVHKWNSYITQYSTGQCTVEMVWELRKENSKRWNPEIQTSNYSKANSANVDTVDNNG 715

QY 583 YTEPRPIGTRYLTP 599  
DB 716 YVEDRLIGTRYLTQNL 732

RESULT 15

Q65444

ID Q65444

AC Q65444;

PRELIMINARY; PRT; 732 AA.

Q65444



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:01:27 ; Search time 13.7812 Seconds  
(without alignments)  
1839.046 Million cell updates/sec

Title: US-09-807-802a-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVQSPQSPDSSS.....NNGLYTEPRPIGTRYLTRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/aaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/aaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/aaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/aaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/aaa/PCUTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/aaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2764.5	85.0	735	US-09-321-589-1	Sequence 1, Appli
2	1830.5	56.3	598	US-09-532-594B-16	Sequence 16, Appli
3	1830.5	56.3	734	US-09-532-594B-4	Sequence 4, Appli
4	1690.5	52.0	544	US-09-532-594B-18	Sequence 18, Appli
5	772	23.7	756	US-09-438-268-4	Sequence 4, Appli
6	479.5	14.7	543	US-08-856-841-22	Sequence 22, Appli
7	326	10.0	415	US-08-856-841-20	Sequence 20, Appli
8	325	10.0	486	US-08-856-841-19	Sequence 19, Appli
9	325	10.0	500	US-08-856-841-16	Sequence 16, Appli
10	325	10.0	501	US-08-856-841-18	Sequence 18, Appli
11	318	9.8	395	US-08-856-841-13	Sequence 13, Appli
12	310	9.5	264	US-08-856-841-14	Sequence 14, Appli
13	303	9.3	298	US-08-856-841-21	Sequence 21, Appli
14	216.5	6.7	579	5223424-13	Patent No. 5223424
15	192	5.9	584	US-09-022-949-2	Sequence 2, Appli
16	178.5	5.5	387	US-08-856-841-17	Sequence 17, Appli
17	119	3.7	1095	US-09-107-532A-3855	Sequence 3855, Ap
18	115	3.5	2736	US-09-252-991A-30227	Sequence 30227, A
19	113.5	3.5	655	US-08-469-202-27	Sequence 27, Appli
20	113.5	3.5	655	US-08-484-434C-34	Sequence 34, Appli
21	113.5	3.5	655	US-09-384-361-34	Sequence 34, Appli
22	112.5	3.5	3060	US-08-487-826B-14	Sequence 14, Appli
23	110.5	3.4	624	US-08-947-965-78	Sequence 78, Appli
24	110.5	3.4	655	US-08-469-202-28	Sequence 28, Appli
25	110.5	3.4	655	US-08-484-434C-35	Sequence 35, Appli
26	110.5	3.4	655	US-09-384-361-35	Sequence 35, Appli
27	109	3.4	1186	US-08-485-568A-4	Sequence 4, Appli

28	109	3.4	1186	1	US-08-357-698-6	Sequence 6, Appli
29	109	3.4	1186	2	US-08-590-554A-4	Sequence 4, Appli
30	109	3.4	1186	2	US-09-184-223-4	Sequence 4, Appli
31	109	3.4	1186	5	PCT-US93-12882-6	Sequence 6, Appli
32	106.5	3.3	1848	3	US-08-296-791-6	Sequence 6, Appli
33	106.5	3.3	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
34	104.5	3.2	717	4	US-09-626-589-1	Sequence 1, Appli
35	104	3.2	1013	3	US-09-415-522-8	Sequence 8, Appli
36	102	3.1	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
37	101.5	3.1	757	4	US-09-252-991A-20231	Sequence 20231, A
38	101	3.1	1651	3	US-09-540-245A-18	Sequence 18, Appli
39	100	3.1	847	4	US-09-373-157-4	Sequence 4, Appli
40	99	3.0	769	4	US-09-252-991A-26777	Sequence 26777, A
41	99	3.0	824	4	US-09-626-589-3	Sequence 3, Appli
42	99	3.0	997	1	US-08-232-540-1	Sequence 1, Appli
43	99	3.0	997	1	US-08-428-949A-1	Sequence 1, Appli
44	99	3.0	997	1	US-08-428-948A-1	Sequence 1, Appli
45	99	3.0	997	2	US-08-428-946-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-321-589-1  
; Sequence 1, Application US/09321589  
; Patent No. 6498244  
; GENERAL INFORMATION:  
; APPLICANT: PATEL, SALIL D.  
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS  
; FILE REFERENCE: 39572  
; CURRENT APPLICATION NUMBER: US/09/321.589  
; CURRENT FILING DATE: 1999-05-28  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: Adeno-associated virus  
US-09-321-589-1

Query Match 85.0%; Score 2764.5; DB 4; Length 735;  
Best Local Similarity 82.8%; Pred. No. 3.3e-237;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY	1	TAPGKKRPVQSPQSPDSSSGIGTKGQPAKRLNFGTGDSESVDPDQPLGEPATPAA	60
DB	138	TAPGKKRFEVHSPVDPDSSSGTGKAGQPAKRLNFGTGDADSVDPDQPLGQPPAASG	197
QY	61	VGPTTASGGGAPMADNNEGADVGNAGNHCDSITWLGDRVITTTSTRTWALPTYNHLY	120
DB	198	LGTNTMATSGAPMADNNEGADVGNAGNHCDSITWMDRVITTTSTRTWALPTYNHLY	257
QY	121	KQISSASTGASNDHYFGYSTPWGYPDFNRPCHFPSPRDWQRLNNNNWGFPRKRLNFKLF	180
DB	258	KQISSQS-GASNDHYFGYSTPWGYPDFNRPCHFPSPRDWQRLNNNNWGFPRKRLNFKLF	316
QY	181	NIQKVEVTNDGVTTIANNLSTVQVFSDSYQIPLYLGSAGHQCCLPPFPADVPMIPOYG	240
DB	317	NIQKVEVTNDGVTTIANNLSTVQVFSYQIPLYLGSAGHQCCLPPFPADVPMIPOYG	376
QY	241	YLTNNGSQVGRSFFCYCLEYFPQMLRTGNFTFSYTFEVPFHSSVAHQSLDRLNMP	300
DB	377	YLTNNGSQVGRSFFCYCLEYFPQMLRTGNFTFSYTFEVPFHSSVAHQSLDRLNMP	436
QY	301	LIDQYLYLNRTQNSQAQNKDLLFRGSGPAGMSVQPKNLPGFCYRQQRVSKTKTDNN	360
DB	437	LIDQYLYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNMLPGFCYRQQRVSKTSADNN	496
QY	361	NSNFTWTCASKYNLNGRESIINPGTAMASHKDDKEDKPFPMGSMVIFGKESAGASNTALDN	420
DB	497	NSEYSWTGATKYHLNGRDSLVPNGPAMASHKDDKEDKFPFQSGVLIFGKQSEKTNVDIEK	556



```
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480
Db 557 VMITDEEIRTNPVATEQGVSTNLQRGRQAATADVNTQGVFGMWQDRDVLQGP 616
QY 481 IWAKIPHTDGHFHPSPLMGFGGLKNPPPOILLKNTVPANPAEASATFASFIQYSTG 540
Db 617 IWAKIPHTDGHFHPSPLMGFGGLKNPPPOILLKNTVPANPSTTSAAKFASFIQYSTG 676
QY 541 QVSVEIWELOKENSKRWNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIGRYLTRPL 599
Db 677 QVSVEIWELOKENSKRWNPEIQTTSYNKSNVNDFTVDNGVISEPREPIGRYLTRNL 735

RESULT 2
US-09-532-594B-16
; Sequence 16, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP2
US-09-532-594B-16

Query Match 56.3%; Score 1830.5; DB 4; Length 598;
Best Local Similarity 57.8%; Pred. No. 3e-154;
Matches .355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;
QY 1 TAPGKRPEVQSPQDPSSSGIGTKTQQPAKRLNF-GQTGDSVSPDPQLGEPPTA 59
Db 1 TAPGKRPLIESPQDPSSSTGIGKKGQPAKKLVFEDETGAAGDPPEGSTG-----A 54
QY 60 AVGPTTMSGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRTVITSTRTWALPTNNHL 119
Db 55 MSDDSEMRRAAGAAVEGGQAGDGVGNAGDWHCDSTWSEGHVTTTSTRTWVLPNTNNHL 114
QY 120 YKQISSASTGASNDNHYGYSTPWGYDFNRPCHFSPRDWORLNNNGWPRPKLNFKL 179
Db 115 YKELGE-----SLQSNYNGFSTPWGYDFNRPCHFSPRDWORLNNNGWMPKAMRVKI 170
QY 180 FNIQKVEVTNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQCLPPFPADVPMI 239
Db 171 FNIQKVEVTNSGETTVANNLTSTVQIFADSSVELPYVMDAGQBSLPPFPNDVPMV 230
QY 240 GY---LTLNNGSAVGRSSFCYLEYFPQSMRLTGNNTFTSYTFFVPHSSVAHSQSLDR 296
Db 231 GYGLVTGNTSQOQTDNRNAPFYCLEYFPQSMRLTGNNTFTSYTFFVPHSSVAHSQSLDR 290
QY 297 LNMPLIDQYLYLNRTQN-----QSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQQRV 352
Db 291 LNMPLIDQYLYLGLQSTTTGTTLNAGTATTN---FTKLRPTNFSNFKWLPGPSIKQGF 347
QY 353 SKTKTDNNNSFTWTGAS---KY-----NLNGRESIINPGTAMASHKDDDEKFPFMSGVMI 405
Db 348 SKTA--NQNYKIPATGSDSLIKYETHSTLDGRWSALTGPPPMATAGPADSK-FSNSQLIF 404
QY 406 FGKESAGASNTALDNWITDEEBIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGAL 465
```

```
Db 405 AGPKQNGTATVPGLTIFTSBEELAAATNATDTMKGMLPGDQSNLPTVDRLTALQAV 464
QY 466 PGMWQDRDVLQGPPIWAKIPHTDGHFHPSPLMGFGGLKNPPPOILLKNTVPANPAEAF 525
Db 465 PGMWQNRDIYYQGPPIWAKIPHTDGHFHPSPLIIGLGLKHPPPQIFIKNTVPANPATTF 524
QY 526 SATKFASFIQYSTGQVSVEIWELOKENSKRWNPEVQYTSNYAKSANVDFTVNNGLYT 585
Db 525 SSTPNSFIQYSTGQVSQVQIDWEIQKRSKRWNPEVQFTSNYQGNLSLLWAPDAAGKYT 584
QY 586 EPRPIGRYLTRPL 599
Db 585 EPRAIGTRYLTHHL 598

RESULT 3
US-09-532-594B-4
; Sequence 4, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP1
US-09-532-594B-4

Query Match 56.3%; Score 1830.5; DB 4; Length 734;
Best Local Similarity 57.8%; Pred. No. 4.2e-154;
Matches 355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;
QY 1 TAPGKRPEVQSPQDPSSSGIGTKTQQPAKRLNF-GQTGDSVSPDPQLGEPPTA 59
Db 137 TAPGKRPLIESPQDPSSSTGIGKKGQPAKKLVFEDETGAAGDPPEGSTG-----A 190
QY 60 AVGPTTMSGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRTVITSTRTWALPTNNHL 119
Db 191 MSDDSEMRRAAGAAVEGGQAGDGVGNAGDWHCDSTWSEGHVTTTSTRTWVLPNTNNHL 250
QY 120 YKQISSASTGASNDNHYGYSTPWGYDFNRPCHFSPRDWORLNNNGWPRPKLNFKL 179
Db 251 YKELGE-----SLQSNYNGFSTPWGYDFNRPCHFSPRDWORLNNNGWMPKAMRVKI 306
QY 180 FNIQKVEVTNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQCLPPFPADVPMI 239
Db 307 FNIQKVEVTNSGETTVANNLTSTVQIFADSSVELPYVMDAGQBSLPPFPNDVPMV 366
QY 240 GY---LTLNNGSAVGRSSFCYLEYFPQSMRLTGNNTFTSYTFFVPHSSVAHSQSLDR 296
Db 367 GYGLVTGNTSQOQTDNRNAPFYCLEYFPQSMRLTGNNTFTSYTFFVPHSSVAHSQSLDR 426
QY 297 LNMPLIDQYLYLNRTQN-----QSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQQRV 352
Db 427 LNMPLIDQYLYLGLQSTTTGTTLNAGTATTN---FTKLRPTNFSNFKWLPGPSIKQGF 483
QY 353 SKTKTDNNNSFTWTGAS---KY-----NLNGRESIINPGTAMASHKDDDEKFPFMSGVMI 405
Db 484 SKTA--NQNYKIPATGSDSLIKYETHSTLDGRWSALTGPPPMATAGPADSK-FSNSQLIF 540
```

```
QY 406 FGKBSAGASNTALDNVMTIDBEEIKATNPVATERGTVAVNPFQSSSTDPATGDVHMGAL 465
Db 541 AGPKQNGTATVPGLTIFTSBEELAAATNATDTDMWGNLPGDQSNLSNLTPTVDRLTALGAV 600
QY 466 PCMWQDRDVLQYQPIWAKIPHTDGHFHPSPLMGGFGLKKNPPQILIKNTVPANPPAEF 525
Db 601 PCMWQNRDIYQYQPIWAKIPHTDGHFHPSPLIIGFGLKHPPPQIFIKNTVPANPATTF 660
QY 526 SATKFPASITQYSTQGVSEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYT 585
Db 661 SSTPVNSFITQYSTQGVSVQIDWEIQKERSKRWNPEVQYTSNYGQNSLLWAPDAAGKYT 720
QY 586 EPRPIGTRYLTRPL 599
Db 721 EPRAIQTRYLTHHL 734

RESULT 4
US-09-532-594B-18
; Sequence 18, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 capsid protein VP3
US-09-532-594B-18

Query Match 52.0%; Score 1690.5; DB 4; Length 544;
Best Local Similarity 59.4%; Pred. No. 7.5e-142;
Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;

QY 67 ASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTVNNHLYKQISSA 126
Db 9 AAGGNAV-EGGQAGDGVGNAGNWHCDSTWSEGHVTTTSTRTWVLTPTVNNHLYKRLGE- 66
QY 127 STGASNDNHYFGYSTPWGYDFNRFCHFPSRDWQRLINNNGVFPKRLNFKLFNIQVKE 186
Db 67 ---SLQSTNYNGFSTPWGYDFNRFCHFPSRDWQRLINNNGVFPKRLNFKLFNIQVKE 123
QY 187 VTTNNGVTTIANNLTSTWQVDSSEVQLPYVLGSAHQGLPFPADVFMIPOGY---LT 243
Db 124 VTTNNGVTTIANNLTSTWQVDSSEVQLPYVLGSAHQGLPFPADVFMIPOGY---LT 183
QY 244 LNNGSQVGRSGFYCYLFPSPQMLRTGNFTPSYTFEEVPHSSVAHSQSLDRLMNPLID 303
Db 184 GNTSQOQTDNFAFYCYLFPSPQMLRTGNFTPSYTFEEVPHSSVAHSQSLDRLMNPLID 243
QY 304 QYLYLNRTON---QSGSAQNKDLLFSRGSFAGMSVQPKWLPFCYRQVRQSKTKTDN 359
Db 244 QYLMGLQSTTTGTTLNAGTATTN---FTKLRPTNFSNFKKNWLPGPSIKQOGFSKTA--N 298
QY 360 NNSNFTWTCAS---KY---NLNGRESIINPGTAMASHKDDDEKPPMSGVMI FGESAG 412
Db 299 QNYKIPATGSDSLIKYETHETSLDGRWSALTPOGPPMAGTAPADSK--FSNSQLIFAGPKQNG 357
QY 413 ASNTALDNVMTIDBEEIKATNPVATERGTVAVNPFQSSSTDPATGDVHMGALPGWVWQD 472
```

```
Db 358 NTATVPOTLLFTSEEEELAAATNATDTDMWGNLPGDQSNLSNLTPTVDRLTALGAVPGWVWQD 417
QY 473 RDVYLQGPWAKIPHTDGHFHPSPLMGGFGLKKNPPQILIKNTVPANPPAEFSAATKFPAS 532
Db 418 RDIYQGPWAKIPHTDGHFHPSPLIIGFGLKHPPPQIFIKNTVPANPATTFSTSTPVNS 477
QY 533 FITQYSTQGVSEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTTPRPITG 592
Db 478 FITQYSTQGVSVQIDWEIQKERSKRWNPEVQYTSNYGQNSLLWAPDAAGKYTTPRAITG 537
QY 593 RYLTRPL 599
Db 538 RYLTHHL 544

RESULT 5
US-09-438-268-4
; Sequence 4, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Virus
US-09-438-268-4

Query Match 23.7%; Score 772; DB 4; Length 756;
Best Local Similarity 32.9%; Pred. No. 6.7e-60;
Matches 195; Conservative 87; Mismatches 257; Indels 54; Gaps 16;

QY 1 TAPGKQKPVESQPOBPSSSGIGTKGQPAKRLNFGQTGDSSEVDPDQPLGEBPATPAA 60
Db 138 TAPGKQKPVESQPOBPSSSGIGTKGQPAKRLNFGQTGDSSEVDPDQPLGEBPATPAA 197
QY 61 VGPTTASGGCAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTVNNHLY 120
Db 198 LGTWTMTSVNSAE-ASTGAGGGGSGNSVKSMWSEGAATFSANSVCTFSQFLIPYDPEHY 256
QY 121 KQISSASTGASND-----NHYFGYSTPWGYDFNRFCHFPSRDWQRLINNNGF 170
Db 257 KVFPAAASCHNASCKEAKVCTISPIMGYSTPWYLDNFALNLFSPLEFQHLIENYCSI 316
QY 171 RPKRLNFKLFNIQVKEVT--TNDGVTTIANNLTSTWQVDSSEVQLPYVLGSAHQGLP 228
Db 317 APDALTVTITISIAVKDVTDKTGGGV-QVTDSTTQRLCMLVDHEYKYVYVLGQGGDTLAP 375
QY 229 FPADVFMIPQGYLTLNN-GSQAVG-----RSFICYLFPSPQMLRTGNFTPSY 278
Db 376 LPVWYFPQVAYLTVGVDVNTQIGSGSKLASSEAFYVLEHSSFQLLTGCTATMYK 435
QY 279 FEEVPHSSVAHSQSLDRLMNPLIDQYLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQ 338
Db 436 PPPVPPENLEGSQHFYEMYNPL---YGSRLGVPDTLGGDPKFRSL-----THEHAIQ 487
QY 339 KNWLPGPCYRQVRQSKTKTDNNSNFTWTCASNYLNGRESIINPG-TAMASHKDDDEK 397
Db 488 QNMFPGPLVNSVSTKEGDSNTGAKALTGLSTGTSTQNTSLRSGFVSPYHHWDTDKY 547
```



STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 415  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PEPTIDE  
DESCRIPTION: N/A  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE: 1  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 20:  
US-08-856-841-20  
Query Match 10.0%; Score 326; DB 3; Length 415;  
Best Local Similarity 26.0%; Pred. No. 1.3e-20;  
Matches 101; Conservative 66; Mismatches 180; Indels 42; Gaps 13;  
193 VTTIANNLSTVQFSDSEYQLPVLSAHQGLPPFPADVPMIPQGYLTNN-GSQAV 251

Db 1 MTMTITPSLHACMLV--DHEYKYPYVLGQGDQTLABELPIWYFFPQYAYLTVDVNTQGI 58  
Qy 252 G-----RSSFCYCLFYFSPQMLRTGNFTFSYTFEEVPHSHSYAHQSILDLWNPFI 302  
Db 59 SGDSKKLASERSAFYVLEHSSFLIGTGCTASMSYKFPVPVPPENLEGCSSQHEFYMYNPL- 117  
Qy 303 DOYLYLARTQNSGSAQNKDLLFSRSGSPAGMSVQPKWLPGPCYRQQRVSKTKTDNNNS 362  
Db 118 --YGSRLGVDPDTLGGDPKFRSL-----THEDHAIQPNFMFGPLVNSVSTKEGDSSTGA 170  
Qy 363 NFWTGTASKYNLNGRESIINFG--TAMASHKDDDEDKFFPMGVMIFGKESAGASNTALDNV 421  
Db 171 GKALTGLSTGTSTQNTSRISLRPGVSPQVYHMDTDKYVTGINAISHGQTTYG--NAEDKE 227  
Qy 422 -----MITDEEIKATNPVATERFGTVAVNFQSSSTDPAFGDVHANGALPGMWQDRD 474  
Db 228 YQGVGRFPNEKEQLQLGLNMHTY-----FPNKGTOYTDQIE-RPLMVGSVNNRA 280  
Qy 475 VYLOQPIWAKIPHTDGHFHS-PLMGFGFLGNPPQIILIKNTVPANPPAFSATKFSAP 533  
Db 281 LHYESQLWSKIPNDDSFQTQPAALGGLGHQPPQIFLK--ILPESGPIGKISMGITT 338  
Qy 534 ITQYSTQVSVIEIWEI--OKENSKRWNP 561  
Db 339 LVQYAVGIMTVMTFKLGRKATGRWNPQ 367

RESULT 8  
US-08-856-841-19  
Sequence 19, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSES: BROOKS HAIDT HAFNER & DELAHUNTY  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:



```
/
/
/ PAGES:
/ DATE: 1982
/ DOCUMENT NUMBER:
/ FILING DATE:
/ PUBLICATION DATE:
/ RELEVANT RESIDUES IN SEQ ID NO: 16:
/ PUBLICATION INFORMATION:
/ AUTHORS: SMITH, D.B.
/ AUTHORS: JOHNSON, K.S.
/ TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
/ TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
/ TITLE: GLUTATHIONE S. TRANSFERASE
/ JOURNAL: GENE
/ VOLUME:
/ ISSUE: 67
/ PAGES: 31 - 40
/ DATE: 1988
/ DOCUMENT NUMBER:
/ FILING DATE:
/ PUBLICATION DATE:
/ RELEVANT RESIDUES IN SEQ ID NO: 16:
/
/ US-08-856-841-16
/
/ Query Match 10.0%; Score 325; DB 3; Length 500;
/ Best Local Similarity 32.2%; Pred. No. 2.1e-20;
/ Matches 88; Conservative 34; Mismatches 127; Indels 24; Gaps 6;
/
/ QY 51 LGEPPATPAAGVPTTMAASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTW 110
/ DB 221 LPEVPAYNASEKYPSTSVNSAE-ASTGAGGGGNSVKSMSEGATFSANVTCTFSRQF 279
/
/ QY 111 ALPTYNNHLYKOISSASTGASND-----NHFGYSTPWGYDFNRFCHSPRDW 160
/ DB 280 LIPYDPEHHYKVPSPAASSCHNAGKEAKVCTISPIWGISTPWRYLDFNALNLFSPLEF 339
/
/ QY 161 ORLNNWGFPRKRLNFKLNFQVKEVT--TNDGVTTIANNLTSTVQVFSSEYOLPYVL 218
/ DB 340 QHLIENYGSIAPDALTVTIASEIAVKDVTDKTGGV-QVTDSTTGRCLMLVDHEYKYPYL 398
/
/ QY 219 GSAHQGCLPPFPADVFMIPOYGYLTNN-GSQAVG-----RSSFYCLEYFSPQMLR 268
/ DB 399 GQGQDLAELPLWVTFPPQYAVLTVDVNTQISGDSKKLASESAFYVLSSHSPQLLG 458
/
/ QY 269 TGNFTFSYTFEEVPPHSSYAHSSQSLDRMLNPL 301
/ DB 459 TGTASMSYKFPVPVENLEGCSQHFYEMYNPL 491
/
/ RESULT 10
/ US-08-856-841-18
/ Sequence 18, Application US/08856841
/ Patent No. 6274307
/ GENERAL INFORMATION:
/ APPLICANT: ERWIN SOUTSCHEK
/ APPLICANT: MANFRED MOTZ
/ TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
/ TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
/ NUMBER OF SEQUENCES: 28
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: BROOKS HAUT HAFNER & DELAHUNTY
/ STREET: 99 PARK AVENUE
/ CITY: NEW YORK
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10016
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" FLOPPY DISC
/ COMPUTER: AT&T - IBM COMPATIBLE
/ OPERATING SYSTEM: MS-DOS Version 6.2
/ SOFTWARE: ASCII
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/856,841
/ FILING DATE:
```

```
/
/
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/214,658
/ FILING DATE: 16-MARCH-1994
/ APPLICATION NUMBER: US 07/917,096
/ FILING DATE: 4-AUGUST-1992
/ APPLICATION NUMBER: PCT/DE91/00106
/ FILING DATE: 8-FEBRUARY-1991
/ APPLICATION NUMBER: DE40038262
/ FILING DATE: 8-FEBRUARY-1990
/ ATTORNEY/AGENT INFORMATION:
/ NAME: ROBINSON, WILLIAM R.
/ REGISTRATION NUMBER: 27,224
/ REFERENCE/DOCKET NUMBER: LKR-9222-A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 697-3355
/ TELEFAX: (212) 557-5635
/ TELEX: NONE
/ INFORMATION FOR SEQ ID NO: 18:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 501
/ TYPE: AMINO ACID
/ TOPOLOGY: LINEAR
/ MOLECULE TYPE:
/ DESCRIPTION: PEPTIDE
/ HYPOTHETICAL: N/A
/ ANTI-SENSE: N/A
/ FRAGMENT TYPE: INTERNAL
/ ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
/ ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)
/ IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
/ POSITION IN GENOME: N/A
/ FEATURE:
/ NAME/KEY: N/A
/ LOCATION: N/A
/ IDENTIFICATION METHOD: amino acid analysis and
/ IDENTIFICATION METHOD: mass spectrometry
/ OTHER INFORMATION:
/ PUBLICATION INFORMATION:
/ AUTHORS: COSSART, Y.E.
/ AUTHORS: FIELD, A.M.
/ AUTHORS: CANT, B.
/ AUTHORS: WIDDOWS, D.
/ TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
/ JOURNAL: LANCET
/ VOLUME: I
/ ISSUE:
/ PAGES: 72 - 73
/ DATE: 1975
/ DOCUMENT NUMBER:
/ FILING DATE:
/ PUBLICATION DATE:
/ RELEVANT RESIDUES IN SEQ ID NO: 18:
/
/ US-08-856-841-18
/
/ Query Match 10.0%; Score 325; DB 3; Length 501;
/ Best Local Similarity 32.2%; Pred. No. 2.1e-20;
/ Matches 88; Conservative 34; Mismatches 127; Indels 24; Gaps 6;
/
/ QY 51 LGEPPATPAAGVPTTMAASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTW 110
/ DB 222 LPEVPAYNASEKYPSTSVNSAE-ASTGAGGGGNSVKSMSEGATFSANVTCTFSRQF 280
/
/ QY 111 ALPTYNNHLYKOISSASTGASND-----NHFGYSTPWGYDFNRFCHSPRDW 160
/ DB 281 LIPYDPEHHYKVPSPAASSCHNAGKEAKVCTISPIWGISTPWRYLDFNALNLFSPLEF 340
/
/ QY 161 ORLNNWGFPRKRLNFKLNFQVKEVT--TNDGVTTIANNLTSTVQVFSSEYOLPYVL 218
/ DB 341 QHLIENYGSIAPDALTVTIASEIAVKDVTDKTGGV-QVTDSTTGRCLMLVDHEYKYPYL 399
/
/ QY 219 GSAHQGCLPPFPADVFMIPOYGYLTNN-GSQAVG-----RSSFYCLEYFSPQMLR 268
/ FILING DATE:
```

Db 400 GQGDTLAPLPIWYFPQYAYLTGVDNTQIGSDSKKLASEBSAFVLEHSSFOLLG 459

Qy 269 TCNNFTFSTEEVPHSSYAHQSGLDRMLNPL 301  
Db 460 TGGTASMSYKFPVPPENLEGCSQHFEYMYNPL 492

## RESULT 11

US-08-856-841-13

; Sequence 13, Application US/08856841

; Patent No. 6274307

; GENERAL INFORMATION:

; APPLICANT: ERWIN SOUTSCHEK

; APPLICANT: MANFRED MOTZ

; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES

; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROOKS HAIDT HAPFNER & DELAHUNTY

; STREET: 99 PARK AVENUE

; CITY: NEW YORK

; STATE: NY

; COUNTRY: USA

; ZIP: 10016

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" FLOPPY DISC

; COMPUTER: AT&T - IBM COMPATIBLE

; OPERATING SYSTEM: MS-DOS Version 6.2

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/856,841

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/214,658

; FILING DATE: 18-MARCH-1994

; APPLICATION NUMBER: US 07/917,096

; FILING DATE: 4-AUGUST-1992

; APPLICATION NUMBER: PCT/DE91/00106

; FILING DATE: 8-FEBRUARY-1991

; APPLICATION NUMBER: DE40038262

; FILING DATE: 8-FEBRUARY-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: ROBINSON, WILLIAM R.

; REGISTRATION NUMBER: 27,224

; REFERENCE/DOCKET NUMBER: LKR-9222-A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 697-3355

; TELEFAX: (212) 557-5635

; TELEX: NONE

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 395

; TYPE: AMINO ACID

; TOPOLOGY: LINEAR

; MOLECULE TYPE: PEPTIDE

; HYPOTHETICAL: N/A

; ANTI-SENSE: N/A

; FRAGMENT TYPE: INTERNAL

; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE

; IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)

; POSITION IN GENOME: N/A

; FEATURE:

; NAME/KEY: N/A

; LOCATION: N/A

; IDENTIFICATION METHOD: amino acid analysis and

; IDENTIFICATION METHOD: mass spectrometry

; OTHER INFORMATION:

; PUBLICATION: COSSART, Y.E.

; AUTHORS: FIELD, A.M.

; AUTHORS: CANT, B.  
; AUTHORS: WIDDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET  
; VOLUME: 1  
; ISSUE:  
; PAGES: 72 - 73  
; DATE: 1975  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 13:  
; PUBLICATION INFORMATION:  
; AUTHORS: MANIATIS, T.  
; AUTHORS: FRITSCH, E.F.  
; AUTHORS: SAMBROOK, J.  
; TITLE: MOLECULAR CLONING  
; JOURNAL: COLD SPRING HARBOR, NY  
; VOLUME:  
; ISSUE:  
; PAGES:  
; DATE: 1982  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 13:  
; PUBLICATION INFORMATION:  
; AUTHORS: SMITH, D.B.  
; AUTHORS: JOHNSON, K.S.  
; TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
; TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
; TITLE: GLUTATHIONE S. TRANSFERASE  
; JOURNAL: GENE  
; VOLUME:  
; ISSUE: 67  
; PAGES: 31 - 40  
; DATE: 1988  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 13:  
; US-08-856-841-13

Query Match 9.8%; Score 318; DB 3; Length 395;

Best Local Similarity 25.4%; Pred. No. 6.1e-20;

Matches 99; Conservative 61; Mismatches 167; Indels 62; Gaps 13;

Qy 193 VTTIANNLSTVQVFSDEYQLPYVLGSAHQGLPPRPADYEMIPQYGYLTNN-GSQAV 251

Db 1 MTMITPSLHACMLV--DHEYKPYVLGGQDTLAPLPIWYFPQYAYLTGVDVNTQGI 58

Qy 252 G-----RSSFYCLEYFPFQMLRTGNFTFSYTFEEVPHSSYAHQSGLDRMLNPLI 302

Db 59 SGDSKCLASEBSAFVLEHSSFQLLGTGTASMSYKFPVPPENLEGCSQHFEYMYNPL- 117

Qy 303 DOYLYLNRTQNGSAQNKDLFSGSPAGMSVQPKMWLPGPCYRQORVSKTKTDNNNS 362

Db 118 --YGSRLGVPTLGGDPKFRSL-----THEDAIQPNFMFGPLVNSVSTKEGSSNTGA 170

Qy 363 NFTWTGASKYNLNGRESIINPG-TAWASHKDEKDFPMSGMVIFGKESAGASNTALDNV 421

Db 171 GKALTGLSTGTSQNTIRSLRFGVSPQYHHWTDKYVTGINAISHGQTYG---NAEDKE 227

Qy 422 -----MITDEEIKATNPVATERFGTVAVNFSSSTDPATGDVHANGALPGMWVQDRD 474

Db 228 YQGVGRFPNEKEQLQLQLNMHTY-----FPNKGTOQVTDQIE-RPLMVGSVMNRA 280

Qy 475 VYVQGPWAKPIHPTDGHFHS-PLMGGEGLKNPPQIILKNTVPANPAFESATKFSF 533

Db 281 LHVESQLSKIPNLDDSFKTQFAALGGWGLHQPFPQI-----P 318

Qy 534 ITQYSTGVSVIEWEL-QKENSKRWNPE 561

Db 319 LKQYAVGIMTVTWTFKLGPKEKATGRWNPQ 347

RESULT 12  
US-08-856-841-14  
; Sequence 14, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: BROOKS HAIDT HAFFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 264  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PEPTIDE  
; DESCRIPTION: N/A  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; ORIGINAL SOURCE: INFECTION (ERTHEMA INFECTIONOSUM)  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
; POSITION IN GENOME: N/A  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.  
; AUTHORS: FIELD, A.M.  
; AUTHORS: CANT, B. D.  
; AUTHORS: WIDDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET

VOLUME: 1  
ISSUE: 1  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES:  
DATE: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
US-08-856-841-14

Query Match 9.5%; Score 310; DB 3; Length 264;  
Best Local Similarity 31.3%; Pred. No. 1.7e-19;  
Matches 78; Conservative 33; Mismatches 104; Indels 34; Gaps 6;

QY 57 TPAAVGPTTMASSGGGAPMADNNEGADGVGNASGNHCHDSTLGLDRVITTTSTRTWALPTYN 116  
Db 5 TFSLAABASTGAGGG-----GNSVKSMSEGATFSANSVCTTFSRQFLIPDP 53  
QY 117 NHLKQISSASTGASND-----NHFGYSTPWGYDFNRRFCHFSPRDWQLINN 166  
Db 54 EHHYKVFSPAASSCHNASGKEAKVCTISPIGYSYTPWRYLDFNALNLPSPLEFQHLIEN 113  
QY 167 NWGFPKRLNFKLFNIQVKEVT--TNDGVTTIANLSTVQVFSSEYQLPYVLGSAHQG 224  
Db 114 YGSIAPDALTVTISEIAVKDVTDKTGGGV-QVTDSTTGLRLCLMLVDHEYKPYVLGGQDT 172  
QY 225 CLPPFPADVEMIPQYGLTLNN-GSOAVG-----RSSFYCLEYFSPSOMLRGTNNPT 274  
Db 173 LAPELPWYFPPQYAYLTVGDVNTQGISGSKKLASAPYVLEHSFQLLGTGCTAS 232  
QY 275 FSYTFEEVP 283  
Db 233 MSYKFPFVP 241

RESULT 13  
US-08-856-841-21  
; Sequence 21, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19



NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HALDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 398  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PEPTIDE  
DESCRIPTION: N/A  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: I  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 21:  
US-08-856-841-21  
Query Match 9.3%; Score 303; DB 3; Length 398;  
Best Local Similarity 25.4%; Pred. No. 1.3e-18;  
Matches 96; Conservative 48; Mismatches 166; Indels 66; Gaps 12;

209 DSEYOLPYVLGSAHQGLPPFPADVFMIPOYGYLTLLNN-GSQAVG-----RSSFYC 258  
4 DHEVKYFVVLGGQODTLAPLPIWVYFPQYAYLTGVDVNTQGISGDSKKLASESAFYV 63  
259 LEYFPPSOMLRGNTNFTSYTFEEVPHSSYAHQSGLDLNPLIDQYLYLNRTQNSGS 318  
64 LEHSSFOLLGTGATATMSYKFPVPPPENLEGCSQHFYEMYNPL-----YGSRLGVDPDTLGGD 120  
319 AQNKDLLFSRGSFAGMSVQPKNLPGPCYRQORVSKTKTDNNNSFNFTWTGASKYNLNGRE 378  
121 PKFRSL-----THEDHAIQONFMPGLVNSVSTKEGSSNTGAKALTGLSTGTSQWTR 175  
379 SIINPG-TAMASHKDDDEKFFPMSGVMI FGKESAGASNTALDNV-----MITDEEIK 430  
176 ISLRPGPVSQPYHHWTDKVTGGINAISHGQTTYG---NAEDKEYQQCGVGRFPPEKEQLK 232  
431 ATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMVQDRDVLVLOGPTWAKIPTHGD 490  
233 QLOGLNMHTY-----FPNKGTOQYTDQIE-RPLMVGSVMNRRLHYESQJWSKIPNLD 285  
491 HFHPS-PLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTQVSVEIEWE 549  
286 SFKTFQFAALGGWGLHQPPI-----FKYHKVQGLEVLNQWE 323  
550 LQENSK-----RWNP 560  
324 LLPFNPWELQLHLNWGP 341  
RESULT 14  
5223424-13  
; Patent No. 5223424  
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,  
; RICHARD D.  
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND  
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO  
; ACID SEQUENCE  
; NUMBER OF SEQUENCES: 16  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/225,032  
; FILING DATE: 27-JUL-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 78,519  
; FILING DATE: 27-JUL-1987  
; APPLICATION NUMBER: 933,107  
; FILING DATE: 20-NOV-1986  
; APPLICATION NUMBER: 902,887  
; FILING DATE: 02-SEP-1986  
; APPLICATION NUMBER: 887,140  
; FILING DATE: 17-JUL-1986  
; APPLICATION NUMBER: 823,102  
; FILING DATE: 27-JAN-1986  
; APPLICATION NUMBER: 773,430  
; FILING DATE: 06-SEP-1985  
; SEQ ID NO: 13:  
; LENGTH: 579  
5223424-13  
Query Match 6.7%; Score 216.5; DB 6; Length 579;  
Best Local Similarity 22.6%; Pred. No. 1.2e-10;  
Matches 143; Conservative 84; Mismatches 250; Indels 157; Gaps 33;  
42 SESVPPDQPLGEPPATPAAGTTPWASGGGAPMANNEGADGVGNASGNWH--CDSWLG 99  
2 SENVEQHPINAGTEL-SATGNESGGGGG---GGRGAGGVGVSTGSFNNQTFEYILG 56  
100 D---RVITSTRTWALPTYNHLYKQIS--SASTGAS---NDNHVFGYSTPWGYDFNR 150  
57 EGLVRIHAHASRLIHLNMPEHETKYRIHVLNSESAGVAGQMVQDDAHTMTVSLIDRNA 116  
151 FHCFSRDPQORLNNMNGFRPKRLNFKLFNTQVKEVT---TNDGVTTIANLNTSTVQVF 207

Db 117 WGVWFPNADQLINNNMTINLVSFEQEIFNVVLKTTITBESAPPSKIYNNDLTASLMVA 176  
Qy 208 SDEYQVLPVLSAHSAGCLPPPADVFMIPQYGL-----TLNNGSAVGRS--- 254  
Db 177 LUTNNTLPYTPAARSETLGFYFWLPKTKQRYLSCIRNLNPPYTGQSQQITDSIQT 236  
Qy 255 -----SFYCLE-YFPQMLRTGNPF-TFSYTFEEVPHSSVAHSQSGLRLMNLIDQYL 306  
Db 237 GLHSDIMFYTIENAVPIHLRLTGDEFSTGIYHFDTKPL--KLTHSWQTNR----- 284  
Qy 307 YYLNRTONSGSAQNKKLLPSRGSPAGMSVQP-----KNWLPQPCYRQ---QRVSKTK 356  
Db 285 -----SLGLPKVLTEPTTEGQHPGLPGANTRKGYHQTINNSY 324  
Qy 357 TD-----NNNSFTWTGASKYNLNGRESIINP--GTAMASHKDDDKFPFMSG 402  
Db 325 TEATLARPAQGVNTPYMFDVSGGFF-----LTFIVTADTQYYDDE----PAGA 372  
Qy 403 VMIFGKESAGASNTALDNVMTIDDEEIKATNP-----VATERFGTVA-VNFQSSS--- 451  
Db 373 IRFTMGYQHGLTTS-----SOELERYTFNQSKCGRAPKQFNQQAFLNLENTNNGTL 426  
Qy 452 ---TDPATG--DVHANGALP-----GMWQDRDVLQGPWAKIPHTD--GHFHPSPLMGG 500  
Db 427 LPSDPTGGKSNKHFMTNLNTYGLTLALNNTAPVFPNGQIWDKELDTDLKPRLLH---VTAP 483  
Qy 501 FGLK-NPPQILIKNTVPANPAEPSA--TKPASPIQYSTGQVSVEIWELOKENSKR 557  
Db 484 FVCKNPPQOLFVKIAP---NLTDFFNADSPQOPRIITD-SNFWKGLTUTFTAKMRSSNM 539  
Qy 558 WNPVQYTSNYAKSANVDFVDNNGLYTEPRPIG 591  
Db 540 WNPQIQHTT-----TAENIRKYI-PTNIG 562

RESULT 15

US-09-022-949-2  
; Sequence 2, Application US/09022949  
; Patent No. 6187759  
; GENERAL INFORMATION:  
; APPLICANT: Tarpey, Ian  
; TITLE OF INVENTION: Canine parvovirus DNA vaccination  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Akzo No. 6187759el Patent Dept.  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/022,949  
; FILING DATE: 11-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gormley, Mary E.  
; REGISTRATION NUMBER: 34,409  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-948-7400  
; TELEFAX: 301-948-9751  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 584 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-09-022-949-2  
Query Match 5.9%; Score 192; DB 3; Length 584;  
Best Local Similarity 19.9%; Pred. No. 1.8e-08;  
Matches 128; Conservative 100; Mismatches 233; Indels 184; Gaps 32;  
Qy 49 QPLGEPPA--TPAAVGPPTTMASGGGAPMADNNEGADGVGNASGNWH-----CDSTWLG 99  
Db 7 QPDGGQPAVRNERATGSGNGSGGG-----GGSGGVGIGISTGTNNKTEPKFLENGV- 59  
Qy 100 DRVITSTRTWALPTYNHLYKQI-----SSASTGASNDNHFGYSTPGYFPDFNRPH 152  
Db 60 -EITANSRLVHLNMPESENTRRVVVNNLDKTAVNGNMALDDTHAQIVTPMSLV DANAWG 118  
Qy 153 CHFSPRDMORLINNNWGPFRKRLNFKLFNIOVKET---TNDGVTTIANNTLTSTVOVPSD 209  
Db 119 VWFNPGDQQLIVNWMSELHLVSFEQEIFNVVLKTVSESATQPPKYVNDLTASLMVALD 178  
Qy 210 SEYQLPYVLGSAHQCL-----PPFPA-----DVFMIPQYGYLTLNNGSAVGRS- 254  
Db 179 SNTMTPTTAAARSETLGFYPMKPTIPTPMRYVYFQWDRTLIPSH-----TGTSGTPTNI 232  
Qy 255 -----SFYCLE-YFPQMLRTGNPF-TFSYTFEEVPHSSVAHSQSGLRLMNLID 303  
Db 233 YHGTDPDDVQVYTIENSVPVHLRLTGDEFATGTFEFCCKP--CRLTHTWQTNRALG--LP 288  
Qy 304 QLYVYLNRTQNS-----GSAQNK-----DLLFSRGSAGMSVQPKNWLPG 344  
Db 289 PFLNSLPQSBGTFNGYIGVQDDKRGVGTQNGNTNYITEATIMRPAEYGS-----A 340  
Qy 345 PCYRQQRVSK-----TKTDNNNSNFTWTGASKY---NLNGRESIINPGT--- 385  
Db 341 PYSFEASTQGFKTPIAAGRGGAQTDENQAA---DGDPRYAFGRQHGOKTTTTGETPER 397  
Qy 386 -AMASHKDDDEKFFPMSGVMIFGKESAGASNTALD-NVMTIDDEEIKATNPVATERFGTV 443  
Db 398 FTYIAHQDT-----GRYPEGDWIONFNLPTVNDNVLPTDPIG-----GKT 440  
Qy 444 AVNFOSSTDPATGDVHAMGALPGMVMDRDVYLQGPWAKIPHTDGHFHPSPLMGGFGL 503  
Db 441 GINY--TNI FNTYGLTALNNVP-----PYPNGQIWDKEFDT-----L 478  
Qy 504 K-----NPPQILIKNTVPAN---PPAEFSATKFASTIYQYSTGQVSVEIE 547  
Db 479 KPRLHVNAFPVCQNNCPQQLFVKVAPNLTNBYDPPDASANMSR-----IVTYSDFMWNKGLIV 534  
Qy 548 WELOKENSKRWNPEVQYTSNYAKSANVDFVDNNGLYTEPRPIG 591  
Db 535 FKAKLRASHTWNPIQQMSIN-----IDNQPNYV-PSNIG 567

Search completed: January 21, 2004, 16:08:02  
Job time : 15.7812 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:04:27 ; Search time 26.9213 Seconds  
(without alignments)  
4549.706 Million cell updates/sec

Title: US-09-807-802a-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQPDSSS.....NNGLYTEPRPICTRYLTRPL 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3251	100.0	736	12	US-10-291-583-64
2	3251	100.0	736	12	US-10-423-704A-5
3	3229	99.3	736	12	US-10-291-583-65
4	2815	86.6	736	12	US-10-291-583-71
5	2815	86.6	736	12	US-10-423-704A-6
6	2764.5	85.0	735	12	US-10-291-583-70
7	2764.5	85.0	735	12	US-10-423-704A-4
8	2764.5	85.0	735	14	US-10-038-972A-13
9	2764.5	85.0	735	15	US-10-293-478-1
10	2759.5	84.7	735	12	US-10-038-972A-14
11	2751.5	84.6	735	12	US-10-291-583-67
12	2751.5	84.6	735	12	US-10-291-583-68
13	2749.5	84.6	735	12	US-10-291-583-69
14	2745.5	84.1	735	12	US-10-291-583-66
15	2735	84.1	736	12	US-10-291-583-97

16	2729	83.9	736	12	US-10-291-583-100
17	2729	83.9	736	12	US-10-423-704A-7
18	2726	83.9	736	12	US-10-291-583-96
19	2722	83.7	736	12	US-10-291-583-98
20	2719	83.6	736	12	US-10-291-583-99
21	2715	83.5	738	12	US-10-291-583-85
22	2715	83.5	738	12	US-10-291-583-91
23	2710	83.4	738	12	US-10-291-583-79
24	2710	83.4	738	12	US-10-291-583-81
25	2709.5	83.3	737	12	US-10-291-583-2
26	2709.5	83.3	737	12	US-10-423-704A-8
27	2707	83.3	738	12	US-10-291-583-93
28	2707	83.3	738	12	US-10-291-583-94
29	2703	83.1	728	12	US-10-291-583-102
30	2703	83.1	728	12	US-10-291-583-80
31	2702	83.1	738	12	US-10-291-583-84
32	2701.5	83.1	731	12	US-10-291-583-89
33	2700	83.1	738	12	US-10-291-583-92
34	2699	83.0	738	12	US-10-291-583-82
35	2698	83.0	728	12	US-10-291-583-106
36	2696.5	82.9	735	12	US-10-291-583-112
37	2696.5	82.9	737	12	US-10-291-583-72
38	2696	82.9	728	12	US-10-291-583-107
39	2696	82.9	728	12	US-10-291-583-108
40	2695	82.9	728	12	US-10-291-583-103
41	2692	82.8	738	12	US-10-291-583-83
42	2691.5	82.8	731	12	US-10-291-583-88
43	2689	82.7	728	12	US-10-291-583-105
44	2686	82.6	728	12	US-10-291-583-101
45	2684.5	82.6	729	12	US-10-291-583-110

ALIGNMENTS

RESULT 1

US-10-291-583-64  
; Sequence 64, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (US20030138772A1) Sequences Identifying Adeno-Associated Virus (US20030138772A1) Sequences  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone AAV1  
US-10-291-583-64

Query Match 100.0%; Score 3251; DB 12; Length 736;  
Best Local Similarity 100.0%; Pred. No. 1.2e-283;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAPGKKRPVEQSPQPDSSSGIGTKGQOPAKRLNFGTGDSVPDPQLGEPPTAA 60

Db 138 TAPGKKRPVEQSPQPDSSSGIGTKGQOPAKRLNFGTGDSVPDPQLGEPPTAA 197

QY 61 VGPTTMSGGGAPMADNNEGADVGNHWCDSWLGDRVITTTSTRTWALPTYNHLY 120

Db 198 VGPPTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY 257  
Qy 121 KOISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWORLNNWNGFRKRLNFKLF 180  
Db 258 KOISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWORLNNWNGFRKRLNFKLF 317  
Qy 181 NIQVKEVTTNDGVTIIANNLTSTVQVFSDSYQLPYVLGSAHQGLPPPPADVFMIPOYG 240  
Db 318 NIQVKEVTTNDGVTIIANNLTSTVQVFSDSYQLPYVLGSAHQGLPPPPADVFMIPOYG 377  
Qy 241 YLTLLNGSAQVGRSSFYCYLFFPSQMLRTGNFTFSYTFEVPFHSSYAHQSQSLDRLMNP 300  
Db 378 YLTLLNGSAQVGRSSFYCYLFFPSQMLRTGNFTFSYTFEVPFHSSYAHQSQSLDRLMNP 437  
Qy 301 LIDQYLYLNRNTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
Db 438 LIDQYLYLNRNTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 497  
Qy 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMI FGKESAGASNTALDN 420  
Db 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMI FGKESAGASNTALDN 557  
Qy 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMWQDRDVLQGP 480  
Db 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMWQDRDVLQGP 617  
Qy 481 IWAKIPHTDGHFSPMLGGFGLKNPPQIILKNTVPANPPAEFSATKFSFITYSTG 540  
Db 618 IWAKIPHTDGHFSPMLGGFGLKNPPQIILKNTVPANPPAEFSATKFSFITYSTG 677  
Qy 541 QVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
Db 678 QVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

## RESULT 2

US-10-423-704A-5  
; Sequence 5, Application US/10423704A  
; Publication No. US20030228282A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvirra, Mauricio  
; TITLE OF INVENTION: Adeno-Associated Virus (AAV) Serotype 8 Sequences, Vectors  
; TITLE OF INVENTION: Containing Same, and Uses Therefor  
; FILE REFERENCE: UPN-02733AUSA  
; CURRENT APPLICATION NUMBER: US/10/423,704A  
; CURRENT FILING DATE: 2003-04-25  
; PRIOR APPLICATION NUMBER: US 60/341,151  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,133  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,122  
; PRIOR FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: PCT/US02/33630  
; PRIOR FILING DATE: 2002-11-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5  
; TYPE: PRT  
; LENGTH: 736  
; ORGANISM: adeno-associated virus serotype 1  
US-10-423-704A-5

Query Match 100.0%; Score 3251; DB 12; Length 736;  
Best Local Similarity 100.0%; Pred. No. 1.2e-283;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TAPGKKRPVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPDQPLGPPATPAA 60  
Db 138 TAPGKKRPVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPDQPLGPPATPAA 197

Qy 61 VGPPTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY 120  
Db 198 VGPPTMASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY 257  
Qy 121 KOISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWORLNNWNGFRKRLNFKLF 180  
Db 258 KOISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWORLNNWNGFRKRLNFKLF 317  
Qy 181 NIQVKEVTTNDGVTIIANNLTSTVQVFSDSYQLPYVLGSAHQGLPPPPADVFMIPOYG 240  
Db 318 NIQVKEVTTNDGVTIIANNLTSTVQVFSDSYQLPYVLGSAHQGLPPPPADVFMIPOYG 377  
Qy 241 YLTLLNGSAQVGRSSFYCYLFFPSQMLRTGNFTFSYTFEVPFHSSYAHQSQSLDRLMNP 300  
Db 378 YLTLLNGSAQVGRSSFYCYLFFPSQMLRTGNFTFSYTFEVPFHSSYAHQSQSLDRLMNP 437  
Qy 301 LIDQYLYLNRNTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
Db 438 LIDQYLYLNRNTQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 497  
Qy 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMI FGKESAGASNTALDN 420  
Db 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMI FGKESAGASNTALDN 557  
Qy 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMWQDRDVLQGP 480  
Db 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMWQDRDVLQGP 617  
Qy 481 IWAKIPHTDGHFSPMLGGFGLKNPPQIILKNTVPANPPAEFSATKFSFITYSTG 540  
Db 618 IWAKIPHTDGHFSPMLGGFGLKNPPQIILKNTVPANPPAEFSATKFSFITYSTG 677  
Qy 541 QVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
Db 678 QVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

## RESULT 3

US-10-291-583-65  
; Sequence 65, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvirra, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (A  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1e1 Sequences Identifi  
; FILE REFERENCE: UPN-02733USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 65  
; TYPE: PRT  
; LENGTH: 736  
; ORGANISM: capsid protein of AAV serotype, clone AAV6VP1  
US-10-291-583-65

Query Match 99.3%; Score 3229; DB 12; Length 736;  
Best Local Similarity 99.2%; Pred. No. 1.1e-281;  
Matches 594; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TAPGKKRPVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPDQPLGPPATPAA 60  
Db 138 TAPGKKRPVEQSPQEPDSSSGIGTKTQQAQKRLNFGQTDGSESVDPDQPLGPPATPAA 197

QY 61 VGPPTWASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYNHLY 120  
DB 198 VGPPTWASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPWGDFDNRFCHFSFPRDQRLINNNWGFPRKRLNFKLF 180  
DB 258 KOISSASTGASNDNHFGYSTPWGDFDNRFCHFSFPRDQRLINNNWGFPRKRLNFKLF 317  
QY 181 NIOQKEVTNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPPADVFMIPQYG 240  
DB 318 NIOQKEVTNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPPADVFMIPQYG 377  
QY 241 YLTLLNGSOAVGRSSFYCLYEPPSOMLRTGNNTFSYTFEEVPHSSYAHQSQSLDLMLNP 300  
DB 378 YLTLLNGSOAVGRSSFYCLYEPPSOMLRTGNNTFSYTFEEVPHSSYAHQSQSLDLMLNP 437  
QY 301 LIDQYLYLNRTOQSGSAQNKDLSFRSGPAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 360  
DB 438 LIDQYLYLNRTOQSGSAQNKDLSFRSGPAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 497  
QY 361 NSNFTWTGASKYKLNNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIIFGKESAGASNTALDN 420  
DB 498 NSNFTWTGASKYKLNNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIIFGKESAGASNTALDN 557  
QY 421 VMTDDEEIKATNPVATEREGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
DB 558 VMTDDEEIKATNPVATEREGTVAVNLQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 617  
QY 481 IWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYSTG 540  
DB 618 IWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYSTG 677  
QY 541 QVSVEIEWELOKENSKRWNEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTIRYLTRPL 599  
DB 678 QVSVEIEWELOKENSKRWNEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTIRYLTRPL 736

RESULT 4  
US-10-291-583--71  
; Sequence 71, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; FILE REFERENCE: US/10291583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 71  
; LENGTH: 736  
; TYPE: PRP  
; ORGANISM: capsid protein of AAV serotype, clone AAV3  
US-10-291-583--71

Query Match 86.6%; Score 2815; DB 12; Length 736;  
Best Local Similarity 85.2%; Pred. No. 2.2e-244;  
Matches 511; Conservative 35; Mismatches 52; Indels 2; Gaps 2;  
QY 1 TAGKCRKRVQSPQBPSSSGIGKGTQGPAPKRLNFGQTDGSDSEVPDPQLGEPPTAPAA 60

DB 138 TAGKKGAVDQSPQBPSSSGIGKGTQGPAPKRLNFGQTDGSDSEVPDPQLGEPPTAPAA 197  
QY 61 VGPPTWASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYNHLY 120  
DB 198 VGPPTWASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPWGDFDNRFCHFSFPRDQRLINNNWGFPRKRLNFKLF 180  
DB 258 KOISSASTGASNDNHFGYSTPWGDFDNRFCHFSFPRDQRLINNNWGFPRKRLNFKLF 316  
QY 181 NIOQKEVTNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPPADVFMIPQYG 240  
DB 317 NIOQKEVTNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPPADVFMIPQYG 376  
QY 241 YLTLLNGSOAVGRSSFYCLYEPPSOMLRTGNNTFSYTFEEVPHSSYAHQSQSLDLMLNP 300  
DB 377 YLTLLNGSOAVGRSSFYCLYEPPSOMLRTGNNTFSYTFEEVPHSSYAHQSQSLDLMLNP 436  
QY 301 LIDQYLYLNRTOQSGSAQNKDLSFRSGPAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 359  
DB 437 LIDQYLYLNRTOQSGSAQNKDLSFRSGPAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 496  
QY 360 NSNFTWTGASKYKLNNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIIFGKESAGASNTALDN 419  
DB 497 NSNFTWTGASKYKLNNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIIFGKESAGASNTALDN 556  
QY 420 NMTDDEEIKATNPVATEREGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 479  
DB 557 NMTDDEEIKATNPVATEREGTVAVNLQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 616  
QY 480 IWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYST 539  
DB 617 IWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYST 676  
QY 540 QVSVEIEWELOKENSKRWNEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTIRYLTRPL 599  
DB 677 QVSVEIEWELOKENSKRWNEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTIRYLTRPL 736

RESULT 5  
US-10-423-704A-6  
; Sequence 6, Application US/10423704A  
; Publication No. US20030228282A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: Adeno-Associated Virus (AAV) Serotype 8 Sequences, Vectors  
; FILE REFERENCE: US/10423704A  
; CURRENT FILING DATE: 2003-04-25  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/341,151  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/377,133  
; PRIOR FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: US 60/386,122  
; PRIOR FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: PCT/US02/33630  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 6  
; LENGTH: 736  
; TYPE: PRP  
; ORGANISM: adeno-associated virus serotype 3  
US-10-423-704A-6

Query Match 86.6%; Score 2815; DB 12; Length 736;  
Best Local Similarity 85.2%; Pred. No. 2.2e-244;  
Matches 511; Conservative 35; Mismatches 52; Indels 2; Gaps 2;  
QY 1 TAGKCRKRVQSPQBPSSSGIGKGTQGPAPKRLNFGQTDGSDSEVPDPQLGEPPTAPAA 60

Db 138 TAPGKGAVDQSPQBPSSGSGVKGSKQAPKRLNFGQTGDSSEVPDPQLGEPAPAPS 197  
Qy 61 VGPPTMASGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTSRTWALPTYNHLY 120  
Db 198 LGSNTMASGGGAPMADNNEGADGVGNSSGNHCHDSTWLGDRVITTSRTWALPTYNHLY 257  
Qy 121 KQISSASTGASNDNHFGYSTPWGYFDNRFCHFSPRDWORLNNNNGFRPKRLNFKLF 180  
Db 258 KQISSQS-GASNDNHFGYSTPWGYFDNRFCHFSPRDWORLNNNNGFRPKRLNFKLF 316  
Qy 181 NIQKVEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQY 240  
Db 317 NIQVRGVTONDGTITTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQY 376  
Qy 241 YLTUNGSQAVGRSSFYCLEYFPPSOMLRTGNNTFTSYTFEEVPHSSVAHSOSLDRLMNP 300  
Db 377 YLTUNGSQAVGRSSFYCLEYFPPSOMLRTGNNTFTSYTFEEVPHSSVAHSOSLDRLMNP 436  
Qy 301 LIDQLYLNRTO-QNOSGAQNKDILLFSGSPAGMSVOPKMWLPGCYRQQRVSKTKTDN 359  
Db 437 LIDQLYLNRTOGTTTNGTSRLLFSQAGPQSHSLQARNWLPGPCYRQQRVSKTKTDN 496  
Qy 360 NNSFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGANTALD 419  
Db 497 NNSNPPWTAASKYHLNGRDLSLVNPGPAMASHKDDKDFPMSGVMI FGKESAGANTALD 556  
Qy 420 NMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 479  
Db 557 NMITDEEIRTNVATEQYGVANNLQSSNTAPTTGTGVNQGALPGMWQDRDVLQGP 616  
Qy 480 PIWAKIPHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPABFSATKFSFITYST 539  
Db 617 PIWAKIPHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPABFSATKFSFITYST 676  
Qy 540 QVSVEIEWELOKENSKRNPVEOYTSNYAKSANVDFTVDNGLYTEPRPICTRYLTRPL 599  
Db 677 QVSVEIEWELOKENSKRNPVEOYTSNYAKSANVDFTVDNGLYTEPRPICTRYLTRPL 736

## RESULT 6

US-10-291-583-70  
; Sequence 70, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilton, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 70  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone AAV2  
US-10-291-583-70  
Query Match 85.0%; Score 2764.5; DB 12; Length 735;  
Best Local Similarity 82.8%; Pred. No. 7.8e-240;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

Qy 1 TAPGKGPVQSPQBPSSGSGIKTQQAPKRLNFGQTGDSSEVPDPQLGEPAPAPS 60  
Db 138 TAPGKGPVQSPQBPSSGSGIKTQQAPKRLNFGQTGDSSEVPDPQLGEPAPAPS 197  
Qy 61 VGPPTMASGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTSRTWALPTYNHLY 120  
Db 198 LGSNTMASGGGAPMADNNEGADGVGNSSGNHCHDSTWLGDRVITTSRTWALPTYNHLY 257  
Qy 121 KQISSASTGASNDNHFGYSTPWGYFDNRFCHFSPRDWORLNNNNGFRPKRLNFKLF 180  
Db 258 KQISSQS-GASNDNHFGYSTPWGYFDNRFCHFSPRDWORLNNNNGFRPKRLNFKLF 316  
Qy 181 NIQKVEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQY 240  
Db 317 NIQVRGVTONDGTITTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQY 376  
Qy 241 YLTUNGSQAVGRSSFYCLEYFPPSOMLRTGNNTFTSYTFEEVPHSSVAHSOSLDRLMNP 300  
Db 377 YLTUNGSQAVGRSSFYCLEYFPPSOMLRTGNNTFTSYTFEEVPHSSVAHSOSLDRLMNP 436  
Qy 301 LIDQLYLNRTO-QNOSGAQNKDILLFSGSPAGMSVOPKMWLPGCYRQQRVSKTKTDN 360  
Db 437 LIDQLYLNRTOGTTTNGTSRLLFSQAGPQSHSLQARNWLPGPCYRQQRVSKTKTDN 496  
Qy 361 NNSFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGANTALD 420  
Db 497 NNSNPPWTAASKYHLNGRDLSLVNPGPAMASHKDDKDFPMSGVMI FGKESAGANTALD 556  
Qy 421 NMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
Db 557 NMITDEEIRTNVATEQYGVANNLQSSNTAPTTGTGVNQGALPGMWQDRDVLQGP 616  
Qy 481 PIWAKIPHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPABFSATKFSFITYST 540  
Db 617 PIWAKIPHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPABFSATKFSFITYST 676  
Qy 541 QVSVEIEWELOKENSKRNPVEOYTSNYAKSANVDFTVDNGLYTEPRPICTRYLTRPL 599  
Db 677 QVSVEIEWELOKENSKRNPVEOYTSNYAKSANVDFTVDNGLYTEPRPICTRYLTRPL 735

## RESULT 7

US-10-423-704A-4  
; Sequence 4, Application US/10423704A  
; Publication No. US2003028282A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilton, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: Adeno-Associated Virus (AAV) Serotype 8 Sequences, Vectors  
; FILE REFERENCE: UPN-02733AUSA  
; CURRENT APPLICATION NUMBER: US/10/423,704A  
; CURRENT FILING DATE: 2003-04-25  
; PRIOR APPLICATION NUMBER: US 60/341,151  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,133  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,122  
; PRIOR FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: PCT/US02/33630  
; PRIOR FILING DATE: 2002-11-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 4  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: adeno-associated virus serotype 2  
US-10-423-704A-4  
Query Match 85.0%; Score 2764.5; DB 12; Length 735;  
Best Local Similarity 82.8%; Pred. No. 7.8e-240;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKKRPVEQSPQEPDSSSGIGKTQQQPAKRLNFGQTGDSESVDPDPOPLGEPPATPAA 60  
DB 138 TAPGKKRPVEHSPVEPDSSSGTGKAGQQQPARKRLNFGQTGDADSDPDPOPLGQPPAAPSG 197  
QY 61 VGPTTMASGGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY 120  
DB 198 LGTNTMATGSGAPMADNNEGADVGNSSGNWHCDSTWMDRVTITSTRTWALPTNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPMGYEDENRPHCHSPRDMQRLINNNWGFPRKRLNFKLF 180  
DB 258 KOISSQS-GASNDNHFGYSTPMGYEDENRPHCHSPRDMQRLINNNWGFPRKRLNFKLF 316  
QY 181 NIQKVEVTTNDGVTITIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPQVG 240  
DB 317 LIQVKEVTQNDGVTITIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVFMIPQVG 376  
QY 241 YLTLLNGSQAAGRSPCYCLEYFPQMLRTGNFTFSTYTPFPHSSVAHQSGLDRLMNP 300  
DB 377 YLTLLNGSQAAGRSPCYCLEYFPQMLRTGNFTFSTYTPFPHSSVAHQSGLDRLMNP 436  
QY 301 LIDQVLYLNRITQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
DB 437 LIDQVLYLNRITQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 496  
QY 241 YLTLLNGSQAAGRSPCYCLEYFPQMLRTGNFTFSTYTPFPHSSVAHQSGLDRLMNP 300  
DB 377 YLTLLNGSQAAGRSPCYCLEYFPQMLRTGNFTFSTYTPFPHSSVAHQSGLDRLMNP 436  
QY 301 LIDQVLYLNRITQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
DB 437 LIDQVLYLNRITQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 496  
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 420  
DB 497 NSEYSWTGATKYLNGRDSLNVNPGPAMASHKDDKDFPMSGVLI FGKQSEKTNVDIEK 556  
QY 421 VMITDEEBEIKATNPATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
DB 557 VMITDEEBEIRTNPVATEQYGSVSTNLQRNQAATAVDNTQGVLPGMVQDRDVLQGP 616  
QY 481 IWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPPAEFSATKFAFITYSTG 540  
DB 617 IWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPPAEFSATKFAFITYSTG 676  
QY 541 QVSVEIEWELQKENS KRWNPVEQYTSNYAKSANVDFTVDDNGLYTEPRPIGTRYLTRPL 599  
DB 677 QVSVEIEWELQKENS KRWNPVEQYTSNYAKSANVDFTVDDNGLYTEPRPIGTRYLTRPL 735

## RESULT 8

US-10-038-972A-13  
; Sequence 13, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038,972A  
; PRIOR FILING DATE: 2002-01-04  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 13  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP1 capsid protein  
US-10-038-972A-13

Query Match 85.0%; Score 2764.5; DB 14; Length 735;  
Best Local Similarity 82.8%; Pred. No. 7.8e-240;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;  
QY 1 TAPGKKRPVEQSPQEPDSSSGIGKTQQQPAKRLNFGQTGDSESVDPDPOPLGEPPATPAA 60  
DB 138 TAPGKKRPVEHSPVEPDSSSGTGKAGQQQPARKRLNFGQTGDADSDPDPOPLGQPPAAPSG 197  
QY 61 VGPTTMASGGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY 120  
DB 198 LGTNTMATGSGAPMADNNEGADVGNSSGNWHCDSTWMDRVTITSTRTWALPTNNHLY 257

QY 121 KOISSASTGASNDNHFGYSTPMGYEDENRPHCHSPRDMQRLINNNWGFPRKRLNFKLF 180  
DB 258 KOISSQS-GASNDNHFGYSTPMGYEDENRPHCHSPRDMQRLINNNWGFPRKRLNFKLF 316  
QY 181 NIQKVEVTTNDGVTITIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPQVG 240  
DB 317 LIQVKEVTQNDGVTITIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVFMIPQVG 376  
QY 241 YLTLLNGSQAAGRSPCYCLEYFPQMLRTGNFTFSTYTPFPHSSVAHQSGLDRLMNP 300  
DB 377 YLTLLNGSQAAGRSPCYCLEYFPQMLRTGNFTFSTYTPFPHSSVAHQSGLDRLMNP 436  
QY 301 LIDQVLYLNRITQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
DB 437 LIDQVLYLNRITQNSGSAQNKDLLFSRGSAGMSVQPKNWLPGCYRQORVSKTKTDNN 496  
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 420  
DB 497 NSEYSWTGATKYLNGRDSLNVNPGPAMASHKDDKDFPMSGVLI FGKQSEKTNVDIEK 556  
QY 421 VMITDEEBEIKATNPATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
DB 557 VMITDEEBEIRTNPVATEQYGSVSTNLQRNQAATAVDNTQGVLPGMVQDRDVLQGP 616  
QY 481 IWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPPAEFSATKFAFITYSTG 540  
DB 617 IWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPPAEFSATKFAFITYSTG 676  
QY 541 QVSVEIEWELQKENS KRWNPVEQYTSNYAKSANVDFTVDDNGLYTEPRPIGTRYLTRPL 599  
DB 677 QVSVEIEWELQKENS KRWNPVEQYTSNYAKSANVDFTVDDNGLYTEPRPIGTRYLTRPL 735

## RESULT 9

US-10-293-478-1  
; Sequence 1, Application US/10293478  
; Publication No. US20030078411A1  
; GENERAL INFORMATION:  
; APPLICANT: PATEL, SALIL D.  
; APPLICANT: MCARTHUR, JAMES G.  
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS  
; FILE REFERENCE: 39672  
; CURRENT APPLICATION NUMBER: US/10/293,478  
; PRIOR FILING DATE: 2002-11-14  
; PRIOR FILING DATE: 1999-05-28  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: Adeno-associated virus  
US-10-293-478-1

Query Match 85.0%; Score 2764.5; DB 15; Length 735;  
Best Local Similarity 82.8%; Pred. No. 7.8e-240;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKKRPVEQSPQEPDSSSGIGKTQQQPAKRLNFGQTGDSESVDPDPOPLGEPPATPAA 60  
DB 138 TAPGKKRPVEHSPVEPDSSSGTGKAGQQQPARKRLNFGQTGDADSDPDPOPLGQPPAAPSG 197  
QY 61 VGPTTMASGGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITTTSTRTWALPTNNHLY 120  
DB 198 LGTNTMATGSGAPMADNNEGADVGNSSGNWHCDSTWMDRVTITSTRTWALPTNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPMGYEDENRPHCHSPRDMQRLINNNWGFPRKRLNFKLF 180  
DB 258 KOISSQS-GASNDNHFGYSTPMGYEDENRPHCHSPRDMQRLINNNWGFPRKRLNFKLF 316  
QY 181 NIQKVEVTTNDGVTITIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPQVG 240

Db 317 NIQVKEVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPOYG 376  
Qy 241 YLTLLNGSAQGRSSFFCYCLEYFPQMLRTGNNTFTSYTFEEVPHSSVAHSQSGLDRLMNP 300  
Db 377 YLTLLNGSAQGRSSFFCYCLEYFPQMLRTGNNTFTSYTFEDVPFHSSVAHSQSGLDRLMNP 436  
Qy 301 LIDQLYLNLRTQNGSQAQNKDLFSGRSPAGMSVQPKWMLPGPCYQQRVSKTKTDNN 360  
Db 437 LIDQLYLNLRTNTPSGTTTQSRQLQFSQAGASDIRDQSRNMLPGPCYQQRVSKTSADNN 496  
Qy 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSGVMIFFKESAGASNTALDN 420  
Db 497 NSEYWTGATKYHLNGRSLVNPAGPAMASHKDDDEKFFPQSGVLLFGQSGSEKTNVDIEK 556  
Qy 421 VMIIDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHMGALPGMWQDRDVLQGP 480  
Db 557 VMIIDEEIRTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMWQDRDVLQGP 616  
Qy 481 IWAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKNTVPANPPAEPFSAFKFASFIQYSTG 540  
Db 617 IWAKIPHTDGHFHPSPMLMGFGGLKHPPQILIKNTVPANPSTTFSAAKFASFIQYSTG 676  
Qy 541 QVSVEIEWELQKENSQRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
Db 677 QVSVEIEWELQKENSQRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 735

RESULT 10  
US-10-038-972A-14  
; Sequence 14, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038, 972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260,124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 598  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP2 capsid protien  
US-10-038-972A-14

Query Match 84.9%; Score 2759.5; DB 14; Length 598;  
Best Local Similarity 82.8%; Pred. No. 1.6e-239;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;  
Qy 2 APGKKRPVEOSPOBPSSSGIGTKGQOPAKKRLNFGQTGDSVDPQPLGPEPATPAAV 61  
Db 2 APGKKRPVEHSPVEPDPSSSGTGAGQOPAKKRLNFGQTGDSVDPQPLGQPPAAPSGL 61  
Qy 62 GPTTMASGGGAPMADNNEGADGVGNAGNHCDSWTGLDRVITTTSTRTWALPTYNHLYK 121  
Db 62 GTTMTATGSGAPMADNNEGADGVGNSSGNHWCDSWTWMDRVITTTSTRTWALPTYNHLYK 121  
Qy 122 QISSASTGASNDNHFCYSTPWGYPFNRPHCHFSPRDWORLNNNNGFRPKRLNFKLFN 181  
Db 122 QISSQS-GASNDNHFCYSTPWGYPFNRPHCHFSPRDWORLNNNNGFRPKRLNFKLFN 180  
Qy 182 IQVKEVTNDGVTITIANNLSTVQVFSDSYQLPYVLGSAHQGCLPPFPADVFMVPOYG 241  
Db 181 IQVKEVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPOYG 240  
Qy 242 LTLNNGSAQGRSSFFCYCLEYFPQMLRTGNNTFTSYTFEEVPHSSVAHSQSGLDRLMNP 301  
Db 241 LTLNNGSAQGRSSFFCYCLEYFPQMLRTGNNTFTSYTFEDVPFHSSVAHSQSGLDRLMNP 300  
Qy 302 IDQLYLNLRTQNGSQAQNKDLFSGRSPAGMSVQPKWMLPGPCYQQRVSKTKTDNN 361

Db 301 IDQLYLNLRTNTPSGTTTQSRQLQFSQAGASDIRDQSRNMLPGPCYQQRVSKTSADNN 360  
Qy 362 SNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSGVMIFFKESAGASNTALDNV 421  
Db 361 SEYWTGATKYHLNGRSLVNPAGPAMASHKDDDEKFFPQSGVLLFGQSGSEKTNVDIEKV 420  
Qy 422 MITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHMGALPGMWQDRDVLQGP 481  
Db 421 MITDEEIRTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMWQDRDVLQGP 480  
Qy 482 WAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKNTVPANPPAEPFSAFKFASFIQYSTG 541  
Db 481 WAKIPHTDGHFHPSPMLMGFGGLKHPPQILIKNTVPANPSTTFSAAKFASFIQYSTG 540  
Qy 542 VSVIEWELQKENSQRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
Db 541 VSVIEWELQKENSQRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 598

RESULT 11  
US-10-291-583-67  
; Sequence 67, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvira, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (?)  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 67  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone A3.7  
US-10-291-583-67

Query Match 84.7%; Score 2753.5; DB 12; Length 735;  
Best Local Similarity 82.8%; Pred. No. 7.7e-239;  
Matches 496; Conservative 42; Mismatches 60; Indels 1; Gaps 1;  
Qy 1 TAPGKKRPVEOSPOBPSSSGIGTKGQOPAKKRLNFGQTGDSVDPQPLGPEPATPA 60  
Db 138 TAPGKKRIEOSPAPEDSDSSGIGKSGQOPAKKRLNFGQTGDSVDPQPLGPEPATPSG 197  
Qy 61 VGPTTMASGGGAPMADNNEGADGVGNAGNHCDSWTGLDRVITTTSTRTWALPTYNHLY 120  
Db 198 VGSNTMASGGGAPMADNNEGADGVGNSSGNHWCDSWTWMDRVITTTSTRTWALPTYNHLY 257  
Qy 121 KOISSASTGASNDNHFCYSTPWGYPFNRPHCHFSPRDWORLNNNNGFRPKRLNFKLFN 180  
Db 258 KOISSQS-GATNDNHFCYSTPWGYPFNRPHCHFSPRDWORLNNNNGFRPKRLNFKLFN 316  
Qy 181 NIQVKEVTNDGVTITIANNLSTVQVFSDSYQLPYVLGSAHQGCLPPFPADVFMVPOYG 240  
Db 317 NIQVKEVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPOYG 376  
Qy 241 YLTLLNGSAQGRSSFFCYCLEYFPQMLRTGNNTFTSYTFEEVPHSSVAHSQSGLDRLMNP 300  
Db 377 YLTLLNGSAQGRSSFFCYCLEYFPQMLRTGNNTFTSYTFEDVPFHSSVAHSQSGLDRLMNP 436  
Qy 301 LIDQLYLNLRTQNGSQAQNKDLFSGRSPAGMSVQPKWMLPGPCYQQRVSKTKTDNN 360



Db 437 LIDQYLYLNRKTQSGTQSRLOFQAGSSWAQAKNWLPGSPYRQRMSTANDNN 496  
Qy 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKODEDEKFPMSGVMIFGKESAGASNTALDN 420  
Db 497 NSEFAWTAATKYLYLNGRNSLVNPGPPWASHKODEDEKFPMHGNIIFGKQGTGTTNVDIES 556  
Qy 421 VMTDDEEIKATNPVATERFQTVAVNFQSSSTDPAIGDVHAMGALPGMWQDRDVLQGP 480  
Db 557 VLIITDEEIRTNVATEQYGVATNHQSNTTASYGVSQGLPLPGMWQDRDVLQGP 616  
Qy 481 IWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFASFTIQTSTG 540  
Db 617 IWAKTPHTDGHFHPSPLMGFGGLKHPPPQILIKNTVPANPATTTFTPGKFASTIQTSTG 676  
Qy 541 QVSVEIEWELQKENSCKWNPVQVTSYNAKSANVDFTVDNNGLYTEPRPIGTTRYLRL 599  
Db 677 QVSVEIEWELQKENSCKWNPVQVTSYNAKSANVDFTVDNNGLYTEPRPIGTTRYLRL 735

## RESULT 12

US-10-291-583-68  
; Sequence 68, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilton, James M.  
; APPLICANT: Alvirra, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (AAV)  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifying  
; FILE REFERENCE: UPN-0273505A  
; CURRENT APPLICATION NUMBER: US/10/291.583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 68  
; TYPE: PRT  
; LENGTH: 735  
; ORGANISM: capsid protein of AAV serotype, clone A3.4  
US-10-291-583-68

Query Match 84.6%; Score 2751.5; DB 12; Length 735;  
Best Local Similarity 82.6%; Pred. No. 1.2e-238;  
Matches 495; Conservative 43; Mismatches 60; Indels 1; Gaps 1;  
Qy 1 TAPGKRPVQSPQEPDSSSGIGTKGQPAKRLNFGQTGDSVPDQPLGEPPTAA 60  
Db 138 TAPGKRPVQSPQEPDSSSGIGTKGQPAKRLNFGQTGDSVPDQPLGEPPTAA 197  
Qy 61 VGPTTASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTTTRTWTALPTNNHLY 120  
Db 198 VGSNTMASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTTTRTWTALPTNNHLY 257  
Qy 121 KQISSASTGASNDNHFGYSTPWGYDFNRFCHFSPRDQWRLNNWGFPRKRLNPKLF 180  
Db 258 KQISSAS-GATNDNHFGYSTPWGYDFNRFCHFSPRDQWRLNNWGFPRKRLNPKLF 316  
Qy 181 NIQKVEVTTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQCLPPFPADVFMIPOYG 240  
Db 317 NIQKVEVTTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQCLPPFPADVFMIPOYG 376  
Qy 241 YLTLLNGSQAVGRSSFFCYCLEYFPSPQMLRTGNFTTFTFEDVPFHSYAHQSQSLDRMLNP 300  
Db 377 YLTLLNGSQAVGRSSFFCYCLEYFPSPQMLRTGNFTTFTFEDVPFHSYAHQSQSLDRMLNP 436

Qy 301 LIDQYLYLNRKTQSGSQAONKDLFSGSPAGMSVQPKNWLPGPCVROORVSKTKTNN 360  
Db 437 LIDQYLYLNRKTQSGTQSRLOFQAGSSWAQAKNWLPGSPYRQRMSTANDNN 496  
Qy 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKODEDEKFPMSGVMIFGKESAGASNTALDN 420  
Db 497 NSEFAWTAATKYLYLNGRNSLVNPGPPWASHKODEDEKFPMHGNIIFGKQGTGTTNVDIES 556  
Qy 421 VMTDDEEIKATNPVATERFQTVAVNFQSSSTDPAIGDVHAMGALPGMWQDRDVLQGP 480  
Db 557 VLIITDEEIRTNVATEQYGVATNHQSNTTASYGVSQGLPLPGMWQDRDVLQGP 616  
Qy 481 IWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFASFTIQTSTG 540  
Db 617 IWAKTPHTDGHFHPSPLMGFGGLKHPPPQILIKNTVPANPATTTFTPGKFASTIQTSTG 676  
Qy 541 QVSVEIEWELQKENSCKWNPVQVTSYNAKSANVDFTVDNNGLYTEPRPIGTTRYLRL 599  
Db 677 QVSVEIEWELQKENSCKWNPVQVTSYNAKSANVDFTVDNNGLYTEPRPIGTTRYLRL 735

## RESULT 13

US-10-291-583-69  
; Sequence 69, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilton, James M.  
; APPLICANT: Alvirra, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (AAV)  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifying  
; FILE REFERENCE: UPN-0273505A  
; CURRENT APPLICATION NUMBER: US/10/291.583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 69  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone A3.5  
US-10-291-583-69

Query Match 84.6%; Score 2749.5; DB 12; Length 735;  
Best Local Similarity 82.6%; Pred. No. 1.8e-238;  
Matches 495; Conservative 43; Mismatches 60; Indels 1; Gaps 1;  
Qy 1 TAPGKRPVQSPQEPDSSSGIGTKGQPAKRLNFGQTGDSVPDQPLGEPPTAA 60  
Db 138 TAPGKRPVQSPQEPDSSSGIGTKGQPAKRLNFGQTGDSVPDQPLGEPPTAA 197  
Qy 61 VGPTTASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTTTRTWTALPTNNHLY 120  
Db 198 VGSNTMASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTTTRTWTALPTNNHLY 257  
Qy 121 KQISSASTGASNDNHFGYSTPWGYDFNRFCHFSPRDQWRLNNWGFPRKRLNPKLF 180  
Db 258 KQISSAS-GATNDNHFGYSTPWGYDFNRFCHFSPRDQWRLNNWGFPRKRLNPKLF 316  
Qy 181 NIQKVEVTTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQCLPPFPADVFMIPOYG 240  
Db 317 NIQKVEVTTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQCLPPFPADVFMIPOYG 376  
Qy 241 YLTLLNGSQAVGRSSFFCYCLEYFPSPQMLRTGNFTTFTFEDVPFHSYAHQSQSLDRMLNP 300  
Db 377 YLTLLNGSQAVGRSSFFCYCLEYFPSPQMLRTGNFTTFTFEDVPFHSYAHQSQSLDRMLNP 436

Qy	301	LIDQYL	YL	NRTQ	NSGSA	QAKND	LL	FSRG	S	PAGMS	VQ	PKWL	PG	PCY	QRO	QVSK	TKTD	NN	360																																		
Db	437	LIDQYL	YL	SKTQ	TG	SGT	Q	SRLO	FNQ	QPS	MA	QA	KWL	PG	PS	YQ	RO	MSK	TAND	NN	496																																
Qy	361	NSNFT	WG	SK	NL	NGR	SI	IN	PG	TAM	A	SH	KD	DE	K	FP	MS	G	YMI	F	KES	SAG	AS	N	TAL	ON	420																										
Db	497	NSEFA	WT	AT	K	Y	PN	GR	SL	N	PG	P	MA	SH	KD	DE	K	FP	MH	GN	L	I	F	K	Q	T	G	T	T	N	D	IES	556																				
Qy	421	VMITD	EE	B	E	I	K	A	N	P	V	A	T	R	G	F	V	A	N	F	O	S	S	T	D	A	T	G	D	V	H	A	M	G	A	L	P	G	W	W	Q	D	R	V	L	Q	P	480					
Db	557	VMITD	EE	B	E	I	K	A	N	P	V	A	T	R	G	F	V	A	N	F	O	S	S	T	D	A	T	G	D	V	H	A	M	G	A	L	P	G	W	W	Q	D	R	V	L	Q	P	516					
Qy	481	IWAKI	P	H	T	D	G	H	P	H	P	S	L	M	G	G	F	G	L	K	N	P	P	Q	I	L	I	K	N	T	P	A	N	P	P	A	B	F	S	A	T	K	F	A	S	I	T	O	Y	S	T	G	540
Db	617	IWAKI	P	H	T	D	G	H	P	H	P	S	L	M	G	G	F	G	L	K	N	P	P	Q	I	L	I	K	N	T	P	A	N	P	A	T	T	F	T	P	G	K	F	A	S	I	T	O	Y	S	T	G	676
Qy	541	QVSVE	I	E	W	L	O	K	E	N	S	K	R	N	N	P	E	Q	Y	T	S	N	A	K	S	A	N	D	F	T	D	N	N	G	L	I	T	E	P	R	P	I	G	T	R	Y	L	R	P	L	599		
Db	677	QVSVE	I	E	W	L	O	K	E	N	S	K	R	N	N	P	E	Q	Y	T	S	N	A	K	S	A	N	D	F	T	D	N	N	G	L	I	T	E	P	R	P	I	G	T	R	Y	L	R	P	L	735		

RESULT 14  
 US-10-291-583-66  
 ; Sequence 66, Application US/10291583  
 ; Publication No. US20030138772A1  
 ; GENERAL INVENTION:  
 ; APPLICANT: Gao, Guangping  
 ; APPLICANT: Wilson, James M.  
 ; APPLICANT: Alvirra, Mauricio  
 ; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (AAV) Sequences and Isolating No. US20030138772A1  
 ; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1  
 ; FILE REFERENCE: UPN-02735USA  
 ; CURRENT APPLICATION NUMBER: US/10/291,583  
 ; CURRENT FILING DATE: 2002-11-12  
 ; PRIOR APPLICATION NUMBER: US 60/350,607  
 ; PRIOR FILING DATE: 2001-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/341,117  
 ; PRIOR FILING DATE: 2001-12-17  
 ; PRIOR APPLICATION NUMBER: US 60/377,066  
 ; PRIOR FILING DATE: 2002-05-01  
 ; PRIOR APPLICATION NUMBER: US 60/386,675  
 ; PRIOR FILING DATE: 2002-06-05  
 ; NUMBER OF SEQ ID NOS: 120  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 66  
 ; LENGTH: 735  
 ; TYPE: PRT  
 ; ORGANISM: capsid protein of AAV serotype, clone A3.3  
 US-10-291-583-66

Query Match	84.5%	Score	2745.5	DB	12	Length	735
Best Local Similarity	82.5%	Pred. No.	4e-238				
Matches	494	Conservative	43	Mismatches	61	Indels	1
Gaps	1						
Qy	1	TAPGCKPVEGSPQSPSSGIGKTGQOPAKRLNFGQTGDSVPDPQIGPEPPAPAA	60				
Db	138	TAPGCKPIEQSPABPSSGIGKSGQOPAKRLNFGQTGDSVPDPQIGPEPPAPSG	197				
Qy	61	VGPTTMASSGGGAPMADNNEGADGVGNASGNHCHDSTWLGRVITTTSTRTWALPTYNHLY	120				
Db	198	VGSNTMASSGGGAPMADNNEGADGVGNSSGNHCHDSTWMDGRVITTTSTRTWALPTYNHLY	257				
Qy	121	KQISSASTGASDNHYPGYSTPWGYDPNRFCHCFSPRDWQRLINNNWGPFPKRLNFKLF	180				
Db	258	KQISSS-GATNDNHYPGYSTPWGYDPNRFCHCFSPRDWQRLINNNWGPFPKRLNFKLF	316				
Qy	181	NIQKVEVTTNDGVTTIANNLTSTVQVPDSEYQLPYVLGSAHQGLCPFPADVFMIPOYG	240				
Db	317	NIQKVEVTQNDGVTTIANNLTSAVQVFTDSEYQLPYVLGSAHQGLCPFPADVFMIPOYG	376				
Qy	241	YLTLLNGSQAVGRSFFCYCLEYFPQOMLTGNNFTTSYTFEYVPHSSYAHQSQSLDRLMNP	300				

377	YLINNGSQAVGRSSFCYLCFFYFSQMLRTGNNFTSFYTFEDVPFHSSVAHQSQDLRLNWP	433
301	LIDQYLVLNRTQNQSSAQNKOLLFRGSGPAGMSVQPNWLPGPCVTRQQRVSKTKTDNN	360
437	LIDQYLVLNRTQNQSSAQNKOLLFRGSGPAGMSVQPNWLPGPCVTRQQRVSKTKTDNN	496
361	NSNFTMTGASKYLNNGRESIINPGTAMASHKDDKDFPMMSGVMI FGKESAGASALTADN	420
497	NSFPANTAAKYLYNNGRSLVNGPVPVASHKDDKEYFPMHGNLI FGKQGTGTTNVDI ES	556
421	VMITDEEIKATNPVATERFGTVAIVFQSSSTD PATGDVHVMAGALPGMVWDRDYYLQGP	480
557	VLI TDEEII TTPVATEQYCGVATNQHQSNTTAS YGSDVDSQGILPGMVWDRDYYLQGP	616
481	IWAKI PHTDGHFHPSLMGGFGLKNPPQILIKNTPVPANPAPAFSATAKPSAFITQYSTG	540
617	IWAKTPHTDGHFHPSLMGGFGLKNPPQILIKNTPVPANPATFTTCKGFPASFTQYSTG	676
541	QVSVEI EWELOKENS KRNWPEVQYTSNYAKSANVDFTVDDNGLYTEPRPIGTRYLTRPL	599
677	QVSVEI EWELOKENS KRNWPEI OYTSNKNKSNVNVFETVDANGVSEPRPIGTRYLTRNL	735

RESULT 15  
US-10-291-583-97  
; Sequence 97, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alvirra, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1 Sequences  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291.583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/350,607  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/341,117  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/377,066  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: US 60/386,675  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 97  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone 43.25  
US-10-291-583-97

Query Match	84.1%	Score 2735;	DB 12;	Length 736;
Best Local Similarity	82.3%	Pred. No. 3.6e-237;		
Matches 494;	Conservative 44;	Mismatches 60;	Indels 2;	Gaps 2
Qy	1	TAPGKKRPVEOS	PQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSESVPDPQPLGEPSPATPAA	60
Db	138	TAPGKKRPVEOS	PQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSESVPDPQPLGEPSPAA	197
Qy	61	VGPTTMASGGGAPMADNNEGADGVGNAGNKHCDSTWLGDAVITTTSTRTWAUPTVNNHLY	120	
Db	198	LGPNTMASGGGAPMADNNEGADGVGNSSGNHCDSTWLGDRVITTTSTRTWAUPTVNNHLY	257	
Qy	121	KQISSASTGAS - NDNHVFGYSTPMGYFPDNRPFCHFSPRDWQRLNNNMGFRPKRLNFKL	179	
Db	258	KQISNGTSGGSTNDNTYTPGYSTPMGYFPDNRPFCHFSPRDWQRLNNNMGFRPKRLNFKL	317	
Qy	180	FNIQVKEVTNDGVTTIANNTLSTVQVPSDSEYQLPYVLGSAHQCLPPFPADVPMI	239	
Db	318	FNIQVKEVTNTEGTYKIANNTLSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVPMVPOY	377	
Qy	240	GYLTLNNGSQAVGRSSFCYCLEBYPFSQMLURTNNGNTFFSYTFEVPFHSYAHQSQDLRLNM	299	



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 15:59:46 ; Search time 32.5714 Seconds  
(without alignments)  
2602.280 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGAPMADNNEGADGVC.....NNGLYTEPRIGTRYLTRPL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.\*

- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2906	100.0	534	21	AAV71169 Adeno-associated v
2	2906	100.0	599	21	AAV71168 Adeno-associated v
3	2906	100.0	736	21	AAV71167 Adeno-associated v
4	2884	99.2	736	22	AAV59847 AAV6 capsid protei
5	2530	87.1	736	22	AAV59846 AAV3B capsid prote
6	2528	87.0	736	22	AAV59845 AAV3A capsid prote
7	2481.5	85.4	533	22	AAV98976 Adeno-associated v
8	2481.5	85.4	533	22	AAV98976 Adeno-associated v
9	2481.5	85.4	598	22	AAV65793 Adeno-associated v

10	2481.5	85.4	598	22	AAV51509 Adeno-associated v
11	2481.5	85.4	598	23	AAU98975 Adeno-associated v
12	2481.5	85.4	735	22	AAV65792 Adeno-associated v
13	2481.5	85.4	735	22	AAV51508 Adeno-associated v
14	2481.5	85.4	735	22	AAV59844 AAV2 capsid protei
15	2481.5	85.4	735	23	AAU98974 Adeno-associated v
16	2477.5	85.3	533	22	AAV51510 Adeno-associated v
17	2455	84.5	734	22	AAV50326 Adeno-associated v
18	1690.5	58.2	544	19	AAV4 VP3 coat prot
19	1690.5	58.2	544	19	AAV4 VP3 coat prot
20	1690.5	58.2	598	19	AAV4 VP2 coat prot
21	1690.5	58.2	598	24	AAV4 VP2 coat prot
22	1690.5	58.2	734	19	AAV4 VP1 capsid pr
23	1690.5	58.2	734	24	AAV4 VP1 capsid pr
24	1690.5	58.2	734	24	ABG73935 Adeno-associated v
25	1676	57.7	736	18	AAV13000 Duck parvovirus ca
26	1676	57.7	534	16	AAV83386 Duck parvovirus ca
27	1665	57.3	732	16	AAV83385 Duck parvovirus ca
28	1665	57.3	532	21	AAV58162 Duck parvovirus ca
29	1665	57.3	532	23	AAU11407 Adeno-associated v
30	1665	57.3	588	21	AAV58161 Adeno-associated v
31	1665	57.3	724	21	AAV58160 Adeno-associated v
32	1665	57.3	724	23	AAU11405 Adeno-associated v
33	491.5	16.9	756	21	AAV71231 Capsid protein enc
34	490	16.9	554	16	AAV08987 Human parvovirus V
35	490	16.9	554	20	AAV23230 Erythrovirus V9 VP
36	490	16.9	781	16	AAV08986 Human parvovirus V
37	490	16.9	781	20	AAV23227 Erythrovirus V9 VP
38	486	16.7	554	24	ABP57264 Human parvovirus B
39	486	16.7	554	24	ABP57267 Human parvovirus B
40	486	16.7	781	24	ABP57263 Human parvovirus B
41	477	16.4	781	24	ABP57266 Human parvovirus B
42	476.5	16.4	543	12	AAV13405 Parvo virus B19 VP
43	304.5	10.5	264	12	AAV13407 Parvo virus B19 PA
44	295.5	10.2	370	12	AAV13406 Parvo virus B19 PA
45	234.5	8.1	579	13	AAV29079 Porcine Parvovirus

## ALIGNMENTS

RESULT 1  
AAV71169  
ID AAV71169 standard; Protein; 534 AA.  
XX  
AC AAV71169;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 capsid protein VP3.  
XX  
DE Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP3.  
XX  
OS Adeno associated virus serotype 1.  
XX  
PN WO200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
DR WPI; 2000-376571/32.  
DR N-PSDB; AAD00772, AAD00779.  
XX  
PT Novel adeno-associated virus serotype 1 polynucleotide useful for

PT preparation of medicament for delivery of a transgene to a host -

PS Claim 7; Page 99-101; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP3 which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 534 AA;

Query Match 100.0%; Score 2906; DB 21; Length 534;  
Best Local Similarity 100.0%; Pred. No. 9.1e-237;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTTRTWALPTNNHLYKQISS 60  
DB 1 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTTRTWALPTNNHLYKQISS 60  
QY 61 ASTGASNDNHFGYSTPWGYPDFNRFCHFSFRDQRLINNNWGRPKRLNPKLFNIQVK 120  
DB 61 ASTGASNDNHFGYSTPWGYPDFNRFCHFSFRDQRLINNNWGRPKRLNPKLFNIQVK 120  
QY 121 EVTTNDGVTITIANLNTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPFMIPOYGYLTIN 180  
DB 121 EVTTNDGVTITIANLNTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPFMIPOYGYLTIN 180  
QY 181 NSQAVGRSSFCYLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLRLMPLIDQY 240  
DB 181 NSQAVGRSSFCYLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLRLMPLIDQY 240  
QY 241 LYYLNRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300  
DB 241 LYYLNRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDFPFPMGVMIFGKESAGASNTALDNNVMTD 360  
DB 301 WTGASKYNLNGRESIINPGTAMASHKDDKDFPFPMGVMIFGKESAGASNTALDNNVMTD 360  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGPFIWAKI 420  
DB 361 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGPFIWAKI 420  
QY 421 PHTDGHFHPSPLMGGFLGNKPPQILLIKTNPVAPNPAEFSATKPFASFTQYSTQGVSE 480  
DB 421 PHTDGHFHPSPLMGGFLGNKPPQILLIKTNPVAPNPAEFSATKPFASFTQYSTQGVSE 480  
QY 481 IEWELQKENS KRNWPEVQYTSYAKSANDVFTVDNNGLYTEPRIGTRYLTRPL 534  
DB 481 IEWELQKENS KRNWPEVQYTSYAKSANDVFTVDNNGLYTEPRIGTRYLTRPL 534

RESULT 2

AA711168  
ID AA711168 standard; Protein; 599 AA.

XX AC AA711168;

XX AC AA711168;

DT 08-SEP-2000 (first entry)

DE Adeno-associated virus serotype 1 capsid protein VP2.

KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;

KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP2.

OS Adeno associated virus serotype 1.

XX WO200028061-A2.

XX 18-MAY-2000.

XX 02-NOV-1999; 99WO-US25694.

XX 05-NOV-1998; 98US-0107114.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Wilson JM, Xiao W;

XX WPI; 2000-376571/32.

XX N-PSDB; AAD00772; AAD00778.

PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host -

PS Claim 7; Page 93-95; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP2 which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 599 AA;

Query Match 100.0%; Score 2906; DB 21; Length 599;  
Best Local Similarity 100.0%; Pred. No. 1.1e-236;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTTRTWALPTNNHLYKQISS 60

DB 66 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTTTRTWALPTNNHLYKQISS 125

QY 61 ASTGASNDNHFGYSTPWGYPDFNRFCHFSFRDQRLINNNWGRPKRLNPKLFNIQVK 120

DB 126 ASTGASNDNHFGYSTPWGYPDFNRFCHFSFRDQRLINNNWGRPKRLNPKLFNIQVK 185

QY 121 EVTTNDGVTITIANLNTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPFMIPOYGYLTIN 180

DB 186 EVTTNDGVTITIANLNTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPFMIPOYGYLTIN 245

QY 181 NSQAVGRSSFCYLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLRLMPLIDQY 240

DB 246 NSQAVGRSSFCYLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLRLMPLIDQY 305

QY 241 LYYLNRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300

DB 306 LYYLNRTQNSGSAQNKDLFLSRGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 365

QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDFPFPMGVMIFGKESAGASNTALDNNVMTD 360

DB 366 WTGASKYNLNGRESIINPGTAMASHKDDKDFPFPMGVMIFGKESAGASNTALDNNVMTD 425

QY 361 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGPFIWAKI 420

DB 426 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGPFIWAKI 485

QY 421 PHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYSTGQVSVE 480  
|||||  
DB 486 PHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYSTGQVSVE 545  
|||||  
QY 481 IEWELQKNSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 534  
|||||  
DB 546 IEWELQKNSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 599  
|||||

RESULT 3  
ID AAY71167 standard; Protein; 736 AA.  
XX  
AC AAY71167;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 capsid protein VP1.  
XX  
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP1.  
XX  
OS Adeno associated virus serotype 1.  
XX  
PN WO200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
DR WPI; 2000-376571/32.  
DR N-PSDB; AAD00772, AAD00777.  
XX

PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host -  
XX  
PS Claim 7; Page 87-90; 108pp; English.  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP1 which is  
CC useful in the production of recombinant viral vector for gene delivery.  
XX

SQ Sequence 736 AA;  
Query Match 100.0%; Score 2906; DB 21; Length 736;  
Best Local Similarity 100.0%; Pred. No. 1.4e-236;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTTRTWALPTYNHLYKQISS 60  
DB 203 MASGGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTTTRTWALPTYNHLYKQISS 262  
QY 61 ASTGASNDNHFGYSTPWGYFDNFRCHFSRDRQRLINNNGWFRPKRLNFKLFNIQVK 120  
|||||

DB 263 ASTGASNDNHFGYSTPWGYFDNFRCHFSRDRQRLINNNGWFRPKRLNFKLFNIQVK 322  
QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPVLGSAHQGCLPPPADVFMIPQYGYLTIN 180  
|||||  
DB 323 EVTTNDGVTTIANNLTSTVQVFSDEYQLPVLGSAHQGCLPPPADVFMIPQYGYLTIN 382  
|||||  
QY 181 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFTSYTPEEVPFHSSYAHQSQSLDRANPLIDQY 240  
|||||  
DB 383 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFTSYTPEEVPFHSSYAHQSQSLDRANPLIDQY 442  
|||||  
QY 241 LYLNRTONQSGSAQNKDLLFSRGSAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSFT 300  
|||||  
DB 443 LYLNRTONQSGSAQNKDLLFSRGSAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSFT 502  
|||||  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDNVMTD 360  
|||||  
DB 503 WTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDNVMTD 562  
|||||  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVYLOGPIWAKI 420  
|||||  
DB 563 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVYLOGPIWAKI 622  
|||||  
QY 421 PHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYSTGQVSVE 480  
|||||  
DB 623 PHTDGHFHPSPMLGGFGLKNPPQILIKNTVPANPPAEFSATKPFASFIQYSTGQVSVE 682  
|||||  
QY 481 IEWELQKNSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 534  
|||||  
DB 683 IEWELQKNSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRIGTRYLTRPL 736  
|||||

RESULT 4  
AAB59847  
ID AAB59847 standard; Protein; 736 AA.  
XX  
AC AAB59847;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE AAV6 capsid protein VP1.  
XX  
KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KW atherosclerosis; sickle cell anaemia; thalassaemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
XX  
OS Adeno associated virus.  
XX  
FN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
DR WPI; 2001-060164/07.  
XX

Adeno-associated virus serotype 6 and viral vector derived from it for  
gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
syndrome, sickle cell anemia, thalassaemia and diabetes -  
XX  
PS Disclosure; Fig 2; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV6). AAV6  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a

CC particular host cell as the AAV virion containing the AAV vector genome  
 CC can be modified to express a capsid protein of an AAV serotype that  
 CC transduces the selected host cell.

XX Sequence 736 AA;  
 SQ Query Match 99.2%; Score 2884; DB 22; Length 736;  
 Best Local Similarity 99.1%; Pred. No. 1e-234;  
 Matches 529; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYKQISS 60  
 DB 203 MASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYKQISS 262  
 QY 61 ASTGASNDNHFGYSTPGWYDFDNRFCHFSFPRDQRLINNNWGRPRKRLNFKLFNIQVK 120  
 DB 263 ASTGASNDNHFGYSTPGWYDFDNRFCHFSFPRDQRLINNNWGRPRKRLNFKLFNIQVK 322  
 QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPOGYLTIN 180  
 DB 323 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPOGYLTIN 382  
 QY 181 NSQAVGRSSFYCLEYFSSQMLRTGNFTFSYTPFVPHSSYAHSQSLDRLMPLIDQY 240  
 DB 383 NSQAVGRSSFYCLEYFSSQMLRTGNFTFSYTPFVPHSSYAHSQSLDRLMPLIDQY 442  
 QY 241 LYLNRQNSGSAQNKDILFSRGSFAGMSVQPKWLPFCPCYRQQRVSKTKTDNNNSFT 300  
 DB 443 LYLNRQNSGSAQNKDILFSRGSFAGMSVQPKWLPFCPCYRQQRVSKTKTDNNNSFT 502  
 QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDEKFPFMSGVMI FKGESAGASNTALDNVMT 360  
 DB 503 WTGASKYNLNGRESIINPGTAMASHKDDKDEKFPFMSGVMI FKGESAGASNTALDNVMT 562  
 QY 361 BEEIKATNPVATERGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 420  
 DB 563 BEEIKATNPVATERGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 622  
 QY 421 PHTDGHFHPSPLMGGFGLKNPPQILIKNTPVPANPPAEFSATKPFASFIQYSTQGVSE 480  
 DB 623 PHTDGHFHPSPLMGGFGLKNPPQILIKNTPVPANPPAEFSATKPFASFIQYSTQGVSE 682  
 QY 481 IEWELQKENS KRNWPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
 DB 683 IEWELQKENS KRNWPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 5  
 AAB59846  
 ID AAB59846 standard; Protein; 736 AA.  
 XX  
 AC AAB59846;  
 XX  
 DT 28-MAR-2001 (first entry)  
 XX  
 DE AAV3B capsid protein VP1.  
 XX  
 KW AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
 KW atherosclerosis; sickle cell anaemia; thalassaemia;  
 KW blood clotting disorder; diabetes; capsid protein VP1.  
 OS Adeno associated virus.

XX US6156303-A.  
 PN  
 XX  
 PD 05-DEC-2000.  
 XX  
 PP 11-JUN-1997; 97US-0873168.  
 XX  
 PR 11-JUN-1997; 97US-0873168.  
 XX  
 PA (UNIW ) UNIV WASHINGTON.

PI Russell DW, Rutledge EA;  
 DR WPI; 2001-060164/07.  
 XX  
 PT Adeno-associated virus serotype 6 and viral vector derived from it for  
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
 XX syndrome, sickle cell anemia, thalassemia and diabetes  
 XX Disclosure; Fig 2; 50pp; English.

CC The present invention relates to adeno-associated virus serotypes. The  
 CC present sequence is capsid protein VP1 of one such serotype (AAV3B).  
 CC AAV3B can be used to construct AAV viral vectors for use in gene therapy  
 CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
 CC sickle cell anemia, thalassemia, blood clotting disorders and diabetes.  
 CC The AAV viral vectors have increased transduction efficiency of a  
 CC particular host cell as the AAV virion containing the AAV vector genome  
 CC can be modified to express a capsid protein of an AAV serotype that  
 CC transduces the selected host cell.

XX Sequence 736 AA;  
 SQ Query Match 87.1%; Score 2530; DB 22; Length 736;  
 Best Local Similarity 86.0%; Pred. No. 8.6e-205;  
 Matches 460; Conservative 27; Mismatches 46; Indels 2; Gaps 2;

QY 1 MASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYKQISS 60  
 DB 203 MASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYKQISS 262  
 QY 61 ASTGASNDNHFGYSTPGWYDFDNRFCHFSFPRDQRLINNNWGRPRKRLNFKLFNIQVK 120  
 DB 263 QS-GASNDNHFGYSTPGWYDFDNRFCHFSFPRDQRLINNNWGRPRKRLNFKLFNIQVK 321  
 QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPOGYLTIN 180  
 DB 322 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPOGYLTIN 381  
 QY 181 NSQAVGRSSFYCLEYFSSQMLRTGNFTFSYTPFVPHSSYAHSQSLDRLMPLIDQY 240  
 DB 382 NSQAVGRSSFYCLEYFSSQMLRTGNFTFSYTPFVPHSSYAHSQSLDRLMPLIDQY 441  
 QY 241 LYLNRQNSGSAQNKDILFSRGSFAGMSVQPKWLPFCPCYRQQRVSKTKTDNNNSNF 299  
 DB 442 LYLNRQNSGSAQNKDILFSRGSFAGMSVQPKWLPFCPCYRQQRVSKTKTDNNNSNF 501  
 QY 300 TWTGASKYNLNGRESIINPGTAMASHKDDKDEKFPFMSGVMI FKGESAGASNTALDNVMT 359  
 DB 502 PWTAAASKYHLNGRDSLVPNPGMAASHKDDKDEKFPFMSGVMI FKGEGTTASNAELDNVMT 561  
 QY 360 DEEIKATNPVATERGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 419  
 DB 562 DEEIKATNPVATERGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 621  
 QY 420 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTPVPANPPAEFSATKPFASFIQYSTQGVSV 479  
 DB 622 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTPVPANPPAEFSATKPFASFIQYSTQGVSV 681  
 QY 480 IEWELQKENS KRNWPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
 DB 682 IEWELQKENS KRNWPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 6  
 AAB59845  
 ID AAB59845 standard; Protein; 736 AA.  
 XX  
 AC AAB59845;  
 XX  
 DT 28-MAR-2001 (first entry)  
 XX  
 DE AAV3A capsid protein VP1.

AAV3A; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
 atherosclerosis; sickle cell anaemia; thalassaemia;  
 blood clotting disorder; diabetes; capsid protein VP1.  
 Adeno associated virus.  
 US6156303-A.  
 05-DEC-2000.  
 11-JUN-1997; 97US-0873168.  
 11-JUN-1997; 97US-0873168.  
 (UNIV ) UNIV WASHINGTON.  
 Russell DW, Rutledge EA;  
 WPI; 2001-060164/07.  
 Adeno-associated virus serotype 6 and viral vector derived from it for  
 gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
 syndrome, sickle cell anemia, thalassaemia and diabetes  
 Disclosure; Fig 2; 50pp; English.  
 The present invention relates to adeno-associated virus serotypes. The  
 present sequence is capsid protein VP1 of one such serotype (AAV3A).  
 AAV3A can be used to construct AAV viral vectors for use in gene therapy  
 for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
 sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
 The AAV viral vectors have increased transduction efficiency of a  
 particular host cell as the AAV virion containing the AAV vector genome  
 can be modified to express a capsid protein of an AAV serotype that  
 transduces the selected host cell.

Query Match 87.0%; Score 2528; DB 22; Length 736;  
 Best Local Similarity 85.8%; Pred. No. 1.3e-204;  
 Matches 459; Conservative 28; Mismatches 46; Indels 2; Gaps 2;

QY 1 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTWALPTYNHLYKQISS 60  
 DB 203 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTWALPTYNHLYKQISS 262

QY 61 ASTGASNDNHFCYSTPMGYFDENRFCHFSPRDQRLINNNWGFPPKRLNFKLFIQK 120  
 DB 263 QS-GASNDNHFCYSTPMGYFDENRFCHFSPRDQRLINNNWGFPPKRLNFKLFIQK 321

QY 121 EVTTNDGVTTIANNLSTVQVSDSYQLPYVLGSAHQCLPPFPADVFMIPQYGLTLN 180  
 DB 322 GVTQNDGTTIANNLSTVQVSDSYQLPYVLGSAHQCLPPFPADVFMIPQYGLTLN 381

QY 181 NGSQAVGRSFCYLEFPSSOMLTGNNFTSYTFEVPFHSSVAHQSLDRLNPLIDQY 240  
 DB 382 NGSQAVGRSFCYLEFPSSOMLTGNNFTSYTFEVPFHSSVAHQSLDRLNPLIDQY 441

QY 241 LYYLNRTO-NQSGSAQKDLFSGSPAGNSVQPKWLPQCYRQVRQVSKTKTDNNNSP 299  
 DB 442 LYYLNRTOGTTTQNSQLLFQAGPQSQSLQARNWLPQCYRQVRQVSKTKTDNNNSP 501

QY 300 TWGASKYNLANGRESINFGTAMASHKDDKDFPFMSGVMI FGKESAGASNTALDNVMT 359  
 DB 502 PWTAAASKYHLNGRSLVNPQPAWASHKDDKDFPFMHGNLIFGKCTTASNAELDNVMT 561

QY 360 DEBEIKATNPVATERGTVAVNPQSSSTDPATGDVHAMGALPGWVQDRDVIYLGPIWAK 419  
 DB 562 DEBEIRTPNPVATEQGTVAVNPQSSSTDPATGDVHAMGALPGWVQDRDVIYLGPIWAK 621

QY 420 IPTHDCGHFSPLMGGFGLKHPPPQILIKNTVPANPAAEFSAKFSITQYSTGOVSV 479  
 DB 622 IPTHDCGHFSPLMGGFGLKHPPPQIMIKNTVPANPAAEFSAKFSITQYSTGOVSV 681

QY 480 EIEWELOKENSCKRNPEVOYTSNYAKSANVDFTVNDNGLYTEPRPIGTRYLTRPL 534  
 DB 682 EIEWELOKENSCKRNPEIQYTSNYNKSNNVDFTVNDNGVYSEPRPIGTRYLTRNL 736

## RESULT 7

AA65794  
 ID AAG65794 standard; Protein; 533 AA.

AC AAG65794;

DT 11-FEB-2002 (first entry)

XX Adeno-associated virus 2 (AAV-2) major coat protein VP3.

XX Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
 KW inverted terminal repeat; neurotropic; neuroprotective; antianemic; ITR;  
 KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
 KW major coat protein; AAV-2; VP3.

OS Adeno-associated virus 2.

XX WO200168888-A2.

PN 20-SEP-2001.

XX 13-MAR-2001; 2001WO-US07927.

XX 14-MAR-2000; 2000US-189110P.

PA (NEUR-) NEUROLOGIX INC.

XX Xiao W, During MJ;

XX WPI; 2001-596912/67.

DR N-PSDB; AAI66974.

PT Recombinant viral vector useful in improving gene therapy in a subject,  
 and for increasing efficiency of entry into a cell, comprises a  
 PT chimeric capsid having one non-native amino acid sequence and a desired  
 PT transgene

PS Disclosure; Page 51; 53pp; English.

XX The invention provides a recombinant viral vector (RVV) comprising a  
 CC chimeric capsid (I) having at least one non-native amino acid sequence,  
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
 CC or their combination, and a transgene flanked 5' and 3' by inverted  
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
 CC combination. The RVV is useful for improving gene therapy in a subject  
 CC with a disorder, and for increasing the efficiency of entry into a cell,  
 CC which involves producing (I) encapsulating a viral vector, and contacting  
 CC a cell with RVV having (I) such that (I) binds to an attachment site on  
 CC the cell surface and permits the vector to enter the cell efficiently.  
 CC A pharmaceutical composition comprising RVV with (I) containing a  
 CC transgene sequence associated with a disease or a disorder such that  
 CC expression of the transgene would result in amelioration of the disease  
 CC or disorder such as inherited neurological and metabolic diseases e.g.  
 CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polyneuropathy,  
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
 CC retinoblastoma and various types of neoplastic cells which include  
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
 CC sarcomas, leukemias and lymphoma. The present sequence represents the  
 CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences  
 CC are used in the construction of a chimeric vector.

SQ Sequence 533 AA;

Query Match 85.4%; Score 2481.5; DB 22; Length 533;  
 Best Local Similarity 83.3%; Pred. No. 6.8e-201;





KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
 XX major coat protein; AAV-2; VP2.  
 OS Adeno-associated virus 2.  
 PH Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "encoded by ACG"  
 XX WO200168888-A2.  
 XX 20-SEP-2001.  
 XX 13-MAR-2001; 2001WO-US07927.  
 XX 14-MAR-2000; 2000US-189110P.  
 XX (NEUR-) NEUROLOGIX INC.  
 XX Xiao W, During MJ;  
 PI WPI; 2001-596912/67.  
 DR N-PSDB; AAI66974.  
 XX Recombinant viral vector useful in improving gene therapy in a subject,  
 PT and for increasing efficiency of entry into a cell, comprises a  
 PT chimeric capsid having one non-native amino acid sequence and a desired  
 PT transgene -  
 XX Disclosure; Page 51; 53pp; English.  
 CC The invention provides a recombinant viral vector (RVV) comprising a  
 CC chimeric capsid (I) having at least one non-native amino acid sequence,  
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
 CC or their combination, and a transgene flanked 5' and 3' by inverted  
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
 CC combination. The RVV is useful for improving gene therapy in a subject  
 CC with a disorder, and for increasing the efficiency of entry into a cell,  
 CC which involves producing (I) encapsulating a viral vector, and contacting  
 CC a cell with RVV having (I) such that (I) binds to an attachment site on  
 CC the cell surface and permits the vector to enter the cell efficiently.  
 CC A pharmaceutical composition comprising RVV with (I) containing a  
 CC transgene sequence associated with a disease or a disorder such that  
 CC expression of the transgene would result in amelioration of the disease  
 CC or disorder such as inherited neurological and metabolic diseases e.g.  
 CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polyneuropathy,  
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
 CC retinoblastoma and various types of neoplastic cells which include  
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
 CC sarcomas, leukemias and lymphoma. The present sequence represents the  
 CC adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences  
 CC are used in the construction of a chimeric vector.  
 XX Sequence 598 AA;  
 SQ Query Match 85.4%; Score 2481.5; DB 22; Length 598;  
 Best Local Similarity 83.3%; Pred. No. 8e-201;  
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
 QY 1 MASGGAPMADNNEGADGVGNAGNHCDSWTWIGDRVITTTSTRTWALPTYNHLYKQISS 60  
 DB 66 MATGSGAPMADNNEGADGVGNAGNHCDSWTWIGDRVITTTSTRTWALPTYNHLYKQISS 125  
 QY 61 ASTGASNDNHYFGVSTPMWGYDFNRFCHSPDQWRLNNNNGFPRKLNFKLNQV 120  
 DB 126 QS-GASNDNHYFGVSTPMWGYDFNRFCHSPDQWRLNNNNGFPRKLNFKLNQV 184  
 QY 121 EVTTNDGVTTIANNLTSTQVDSSEYQLPYVLGSAHQGLPPFPADVFMPQGYLTN 180  
 DB 185 EVTQNDGTTIANNLTSTQVDSSEYQLPYVLGSAHQGLPPFPADVFMPQGYLTN 244

QY 181 NGSQVGRSSFYCLEYFPFSQMLRTGNFTSYTFBEVPHSHSYAHQSQSLDRLMPLIDQY 240  
 DB 245 NGSQVGRSSFYCLEYFPFSQMLRTGNFTSYTFBEVPHSHSYAHQSQSLDRLMPLIDQY 304  
 QY 241 LYIARTNQSGSAQNKDLLFSRGSFAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 300  
 DB 305 LYLSRTNTTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQORVSKTSADNNSEYS 364  
 QY 301 WTGASKYNLNGRESIINPGTAMASHKDDDEKFPFMSGVMI FCKESAGASANTALDNNMID 360  
 DB 365 WTGATKYHLNGRDSLUNPGPAMASHKDDDEKFPFMSGVMI FCKESAGASANTALDNNMID 424  
 QY 361 EEEIKATNPVATEREGTVAVNFPQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKI 420  
 DB 425 EEEIRTNFVATEQYGVSVSTNLQRGNQAATADVNTQGVLPGMVWQDRDVLQGPWAKI 484  
 QY 421 PHTDGHFHPSPLMGGFGLKNPPQQLIKNTVPANPPAEFSATKFASTFQYSTQGVSV 480  
 DB 485 PHTDGHFHPSPLMGGFGLKNPPQQLIKNTVPANPPSTTFSAKFASFQYSTQGVSV 544  
 QY 481 IEWELQKENSKRWNPEVOYTSNYSKANSVDFTDVNNGLYTEPRPIGTIRYLTRPL 534  
 DB 545 IEWELQKENSKRWNPEIQTYSNYSKANSVDFTDVNTGVYSEPRPIGTIRYLTRNL 598  
 RESULT 10  
 AAM51509  
 ID AAM51509 standard; protein; 598 AA.  
 AC AAM51509;  
 XX  
 DT 02-JAN-2002 (first entry)  
 DE Adeno-associated virus VP2 capsid protein.  
 KW Adeno-associated virus; AAV; VP2; capsid; virus-like particle;  
 KW nuclear localisation signal; VP3.  
 OS Adeno associated virus.  
 XX JP2001169777-A.  
 PD 26-JUN-2001.  
 PF 30-JUL-1999; 99JP-0249140.  
 PR 30-JUL-1999; 99JP-0249140.  
 PA (HAND/) HANDA H.  
 DR WPI; 2001-599854/68.  
 XX New virus-like particles from VP3 capsid protein of adeno-associated  
 PT virus, comprise a peptide containing a nucleus-shifting signal  
 PT connected to its N-terminal -  
 PS Claim 1; Page 14-16; 33pp; Japanese.  
 CC The present sequence is provided in a specification relating to a  
 CC virus-like particle-forming protein, and to a peptide containing a  
 CC nuclear-shifting signal at its N-terminus that can form virus-like  
 CC particles by shifting to the nucleus of the animal cell in which it  
 CC is expressed. The method is used for forming virus-like particles  
 CC from the VP3 capsid protein of adeno-associated virus (AAV).  
 XX Sequence 598 AA;  
 SQ Query Match 85.4%; Score 2481.5; DB 22; Length 598;  
 Best Local Similarity 83.3%; Pred. No. 8e-201;  
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
 QY 1 MASGGAPMADNNEGADGVGNAGNHCDSWTWIGDRVITTTSTRTWALPTYNHLYKQISS 60

Db 66 MATSGAPMADNNEGADGVGNSGNHCHDSTWMDRVITSTRTWALPTYNHLYKQISS 125  
 QY 61 ASTGASNDNHFGYSTPWGYPDFNRFCHFSPRDWORLNNNNNGFRPKLNFLENIOVK 120  
 Db 126 QS-GASNDNHFGYSTPWGYPDFNRFCHFSPRDWORLNNNNNGFRPKLNFLENIOVK 184  
 QY 121 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQYGYLTIN 180  
 Db 185 EVTQNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQYGYLTIN 244  
 QY 181 NSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 240  
 Db 245 NSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 304  
 QY 241 LYLLNRTQNSGSAQNKDLLEFRSGSPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 300  
 Db 305 LYLLSRTNTPSGTTTQSRQLQFQAGASDIRDQSRNWLPGPCYRQORVSKTSADNNNSYS 364  
 QY 301 WTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSGVMI FKGESAGASNTALDNVMTD 360  
 Db 365 WTGATKYHLNGRDSLNVNPGPAMASHKDDDEKFFPQSGVLIFGKQSGSEKTNVDIEKVMITD 424  
 QY 361 EEEIKATNPVATERFGTVAVNFPSSSTDPATGVDHAMGALPGMWQDRDVLQGPWIWAKI 420  
 Db 425 EEEIRTNPVATEQYGVSVTNLQGRNQATADVNTQGVLPQWQDRDVLQGPWIWAKI 484  
 QY 421 PHTDGHFHPSPMLMGFGGLKHPPPQILIKNTVPANPPAFSATKFSFITYSTGQSVSE 480  
 Db 485 PHTDGHFHPSPMLMGFGGLKHPPPQILIKNTVPANPPAFSATKFSFITYSTGQSVSE 544  
 QY 481 IEWELQKENSXRNWPEVQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
 Db 545 IEWELQKENSXRNWPEIQVTSNYKSNVNDFTVDNNGVYSEPRPIGTRYLTRNL 598

## RESULT 11

AAU98975  
 ID AAU98975 standard; Protein; 598 AA.  
 AC AAU98975;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE Adeno-associated virus 2 (AAV2) vector, VP2 capsid protein.  
 XX  
 KW Adeno-associated virus 2 vector; AAV2; cancer; VP2 capsid;  
 KW heparin-sulphate proteoglycan; vaccine; immune response;  
 KW ovarian cancer.  
 XX  
 OS Adeno-associated virus 2.  
 XX  
 PN WO200253703-A2.  
 XX  
 PD 11-JUL-2002.  
 XX  
 PP 04-JAN-2002; 2002WO-US00152.  
 XX  
 PR 05-JAN-2001; 2001US-260124P.  
 XX  
 PA (CHIL-) CHILDRENS HOSPITAL INC.  
 XX  
 PI Bartlett JS;  
 XX  
 DR WPI; 2002-583608/62.  
 DR N-PSDB; ABK89694.  
 XX  
 PT New adeno-associated virus vector comprises a biotinylated capsid or  
 PT capsid protein with an amino acid insertion in the VP1 capsid, useful  
 PT as a vaccine or for transferring a therapeutic peptide to a cancer cell  
 PT .  
 XX  
 PS Disclosure; Page 52-54; 57pp; English.  
 XX

CC The invention relates to an adeno-associated virus (AAV) vector (I)  
 CC comprising a biotinylated capsid or capsid protein (II) with an amino  
 CC acid insertion following the capsid amino acid at position 139, 161, 588  
 CC or 657 in the VP1 capsid. The AAV vector comprises a capsid protein  
 CC containing one or more amino acid insertions that ablate the ability of  
 CC the vector to bind heparin-sulphate proteoglycan and allow the vector to  
 CC use a cellular receptor not used by wild type AAV. Modified (I) are  
 CC useful as vaccines to elicit immune responses to amino acids, where the  
 CC response can be protective and/or therapeutic. (I) may be used to  
 CC transfer a therapeutic peptide to a cancer cell, particularly to an  
 CC ovarian cancer cell. The present sequence represents the adeno-associated  
 CC virus 2 (AAV2) vector, VP2 capsid protein used in to make modified  
 CC AAV2 vectors.  
 XX

## Sequence 598 AA;

Query Match 85.4%; Score 2481.5; DB 23; Length 598;  
 Best Local Similarity 83.3%; Pred. No. 8e-201;  
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
 QY 1 NASGGAPMADNNEGADGVGNSGNHCHDSTWMDRVITSTRTWALPTYNHLYKQISS 60  
 Db 66 MATSGAPMADNNEGADGVGNSGNHCHDSTWMDRVITSTRTWALPTYNHLYKQISS 125  
 QY 61 ASTGASNDNHFGYSTPWGYPDFNRFCHFSPRDWORLNNNNNGFRPKLNFLENIOVK 120  
 Db 126 QS-GASNDNHFGYSTPWGYPDFNRFCHFSPRDWORLNNNNNGFRPKLNFLENIOVK 184  
 QY 121 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQYGYLTIN 180  
 Db 185 EVTQNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPQYGYLTIN 244  
 QY 181 NSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 240  
 Db 245 NSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 304  
 QY 241 LYLLNRTQNSGSAQNKDLLEFRSGSPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 300  
 Db 305 LYLLSRTNTPSGTTTQSRQLQFQAGASDIRDQSRNWLPGPCYRQORVSKTSADNNNSYS 364  
 QY 301 WTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSGVMI FKGESAGASNTALDNVMTD 360  
 Db 365 WTGATKYHLNGRDSLNVNPGPAMASHKDDDEKFFPQSGVLIFGKQSGSEKTNVDIEKVMITD 424  
 QY 361 EEEIKATNPVATERFGTVAVNFPSSSTDPATGVDHAMGALPGMWQDRDVLQGPWIWAKI 420  
 Db 425 EEEIRTNPVATEQYGVSVTNLQGRNQATADVNTQGVLPQWQDRDVLQGPWIWAKI 484  
 QY 421 PHTDGHFHPSPMLMGFGGLKHPPPQILIKNTVPANPPAFSATKFSFITYSTGQSVSE 480  
 Db 485 PHTDGHFHPSPMLMGFGGLKHPPPQILIKNTVPANPPAFSATKFSFITYSTGQSVSE 544  
 QY 481 IEWELQKENSXRNWPEVQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
 Db 545 IEWELQKENSXRNWPEIQVTSNYKSNVNDFTVDNNGVYSEPRPIGTRYLTRNL 598

## RESULT 12

AAAG65792  
 ID AAAG65792 standard; Protein; 735 AA.  
 AC AAAG65792;  
 XX  
 DT 11-FEB-2002 (first entry)  
 XX  
 DE Adeno-associated virus 2 (AAV-2) major coat protein VP1.  
 XX  
 KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
 KW inverted terminal repeat; neurotropic; neuroprotective; antianemic; ITR;  
 KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
 KW major coat protein; AAV-2; VP1.  
 XX  
 OS Adeno-associated virus 2.

XX WO200168888-A2.  
XX 20-SEP-2001.  
XX 13-MAR-2001; 2001WO-US07927.  
XX 14-MAR-2000; 2000US-189110P.  
XX (NEUR-) NEUROLOGIX INC.  
XX Xiao W, During MJ;  
XX WPI; 2001-596912/67.  
XX N-PSDB; AAI66974.  
XX Recombinant viral vector useful in improving gene therapy in a subject,  
XX and for increasing efficiency of entry into a cell, comprises a  
XX chimeric capsid having one non-native amino acid sequence and a desired  
XX transgene  
XX  
XX Disclosure; Page 50; 53pp; English.  
XX The invention provides a recombinant viral vector (RVV) comprising a  
XX chimeric capsid (I) having at least one non-native amino acid sequence,  
XX derived from a capsid protein domain of parvovirus (II), a virus (III),  
XX or their combination, and a transgene flanked 5' and 3' by inverted  
XX terminal repeat (ITR) sequences derived from (II), (III), or their  
XX combination. The RVV is useful for improving gene therapy in a subject  
XX with a disorder, and for increasing the efficiency of entry into a cell,  
XX which involves producing (I) encapsulating a viral vector, and contacting  
XX a cell with RVV having (I) such that (I) binds to an attachment site on  
XX the cell surface and permits the vector to enter the cell efficiently.  
XX A pharmaceutical composition comprising RVV with (I) containing a  
XX transgene sequence associated with a disease or a disorder such that  
XX expression of the transgene would result in amelioration of the disease  
XX or disorder such as inherited neurological and metabolic diseases e.g.  
XX lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polyneuropathy,  
XX Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
XX e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
XX fibrosis, diabetes, diseases associated with hormone deficiencies,  
XX retinoblastoma and various types of neoplastic cells which include  
XX tumours especially central nervous system tumours, neoplasms, carcinomas,  
XX sarcomas, leukemias and lymphoma. The present sequence represents the  
XX adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences  
XX are used in the construction of a chimeric vector.  
XX  
XX Sequence 735 AA;

Query Match 85.4%; Score 2481.5; DB 22; Length 735;  
Best Local Similarity 83.3%; Pred. No. 1.1e-200;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
SQ  
1 MASGGAPADNNEGADGVGNAGNHCDSWLGDRVITTTTRTWTALPTNNHLYKQISS 60  
203 MATGSGAPADNNEGADGVGNAGNHCDSWLGDRVITTTTRTWTALPTNNHLYKQISS 262  
61 ASTGASNDNHFGYSTPMGYDFNRFCHFSFPRDQRLNNNNGWPRKRLNFKLFNIQVK 120  
263 QS-GASNDNHFGYSTPMGYDFNRFCHFSFPRDQRLNNNNGWPRKRLNFKLFNIQVK 321  
121 EVTTNDGVTTIANNLTSTQVFSSEYQLPYVLGSAHQCLPPFPADVFMIQYGYLTIN 180  
322 EVTQNDGVTTIANNLTSTQVFTDSEYQLPYVLGSAHQCLPPFPADVFMIQYGYLTIN 381  
181 NGSQAVGRSFCYLEFPSCMLRTGNFTFTFEEVPFHSSYAHSSQSLDRMLNPLIDQY 240  
382 NGSQAVGRSFCYLEFPSCMLRTGNFTFTFEEVPFHSSYAHSSQSLDRMLNPLIDQY 441  
241 LYTANRQNSQSAQNKLDFSRGSPAGMSVOPKWLPGPCYRQORVSKTKTDNNNSFT 300  
442 LYLRSNTFTSGTTQSRQLQFQAGASDIRDSRNWLPGPCYRQORVSKTSADNNNSYS 501

QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKFFPMGSGYMI FCKESAGASNTALDNVMTD 360  
DB 502 WTGATKYHLNGRDSLIVNPGPAMASHKDDKFFPGSGVLIFGKQSEKTNVDIEKVMITD 561  
QY 361 EEEIKATNPVATERFETVAVNPFQSSSTDPATGDVHAMGALPGVMVQDRDVIYLGQPIWAKI 420  
DB 562 EEEIRTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVQDRDVIYLGQPIWAKI 621  
QY 421 PHTDGHFSPPLMGCGFGLKNPPQILIKNTPVPANPPAEFSATKPFITQYSTQGVSYE 480  
DB 622 PHTDGHFSPPLMGCGFGLKNPPQILIKNTPVPANPPSTTFSAAKPFITQYSTQGVSYE 681  
QY 481 IEWELOKENSKEWNPVEVOYTSNYAKSANVDFTVDNGLYTEPRPIGTRVLTPL 534  
DB 682 IEWELOKENSKEWNPFIQYTSNYKSNVVDFTVDNGVYSEPRPIGTRVLTPL 735

RESULT 13  
AAM51508  
ID AAM51508 standard; protein; 735 AA.  
XX  
XX AAM51508;  
XX  
XX 02-JAN-2002 (first entry)  
XX Adeno-associated virus VP1 capsid protein.  
XX  
XX Adeno-associated virus; AAV; VP1; capsid; virus-like particle;  
XX nuclear localisation signal; VP3.  
XX Adeno associated virus.  
XX JP2001169777-A.  
XX 26-JUN-2001.  
XX 30-JUL-1999; 99JP-0249140.  
XX 30-JUL-1999; 99JP-0249140.  
XX (HAND/) HANDA H.  
XX WPI; 2001-599854/68.  
XX New virus-like particles from VP3 capsid protein of adeno-associated  
XX virus, comprise a peptide containing a nucleus-shifting signal  
XX connected to its N-terminal -  
XX  
XX Disclosure; Page 10-13; 33pp; Japanese.  
XX The present sequence is provided in a specification relating to a  
XX virus-like particle-forming protein, and to a peptide containing a  
XX nuclear-shifting signal at its N-terminus that can form virus-like  
XX particles by shifting to the nucleus of the animal cell in which it  
XX is expressed. The method is used for forming virus-like particles  
XX from the VP3 capsid protein of adeno-associated virus (AAV).  
XX  
XX Sequence 735 AA;

Query Match 85.4%; Score 2481.5; DB 22; Length 735;  
Best Local Similarity 83.3%; Pred. No. 1.1e-200;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
QY 1 MASGGAPADNNEGADGVGNAGNHCDSWLGDRVITTTTRTWTALPTNNHLYKQISS 60  
DB 203 MATGSGAPADNNEGADGVGNAGNHCDSWLGDRVITTTTRTWTALPTNNHLYKQISS 262  
QY 61 ASTGASNDNHFGYSTPMGYDFNRFCHFSFPRDQRLNNNNGWPRKRLNFKLFNIQVK 120  
DB 263 QS-GASNDNHFGYSTPMGYDFNRFCHFSFPRDQRLNNNNGWPRKRLNFKLFNIQVK 321  
QY 121 EVTTNDGVTTIANNLTSTQVFSSEYQLPYVLGSAHQCLPPFPADVFMIQYGYLTIN 180





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:00:42 ; Search time 12.5714 Seconds  
(without alignments)  
4084.983 Million cell updates/sec

Title: US-09-807-802A-17  
Perfect score: 2906  
Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRIGTYRLTRPL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2177.5	74.9	504	1 VCPV3A	coat protein - ade
2	1676	57.7	732	2 S52210	coat protein VP1 -
3	517	17.8	673	1 VCPVB5	coat protein VP1 -
4	487.5	16.8	781	1 VCPV19	coat protein VP1 -
5	234.5	8.1	723	1 VCPVPP	coat protein VP1 -
6	232.5	8.0	729	1 VCPVNA	coat protein VP1 -
7	232.5	8.0	729	1 A60006	coat protein VP1 -
8	229	7.9	587	1 B44276	coat protein VP1 -
9	226	7.8	722	1 VCPV2	coat protein VP1 -
10	197	6.8	716	1 VCPV2M	coat protein VP1 -
11	196.5	6.8	718	1 VCPVIM	coat protein VP1 -
12	194	6.7	722	1 VCPVME	coat protein VP1 -
13	194	6.7	727	1 VCPV1F	coat protein VP1 -
14	189	6.5	584	2 S49594	capaid protein VP2
15	189	6.5	727	1 VCPVFP	coat protein VP1 -
16	188	6.5	737	1 VCPVCD	coat protein VP1 -
17	184	6.3	748	1 VCPVCP	coat protein VP1 -
18	177.5	6.1	702	1 VCPVAP	coat protein VP1 -
19	175	6.0	722	1 VCPVCN	coat protein VP1 -
20	123.5	4.2	690	2 AB0124	probable TonB-depe
21	122.5	4.2	880	1 SYBSVS	valine-tRNA ligase
22	121	4.2	635	2 P95660	protein FK11.10 [
23	118	4.1	739	2 T52289	probable transketo
24	115.5	4.0	345	1 G97024	probable phospho
25	113.5	3.9	2817	2 B97033	uncharacterized pr
26	109	3.8	418	2 T35753	probable periplasm
27	109	3.8	1072	2 A86827	hypothetical prote
28	109	3.8	1186	2 T18210	delta endotoxin -
29	108.5	3.7	1711	2 AB1283	peptidoglycan link

30	108	3.7	655	1 ALKBG	cyclomaltodextrin
31	108	3.7	772	1 A44052	outer layer protei
32	108	3.7	826	2 AD1683	penicillin-binding
33	107.5	3.7	413	2 T23098	hypothetical prote
34	107.5	3.7	648	2 S50856	whn protein - rat
35	107	3.7	777	2 A80462	probable exported
36	107	3.7	1338	2 T30565	MAP kinase kinase
37	106.5	3.7	403	2 B87513	acyl-CoA dehydroge
38	106	3.6	1271	2 D54237	hypothetical prote
39	105.5	3.6	2500	2 G71609	hypothetical prote
40	105	3.6	1394	2 A29637	position-specific
41	104.5	3.6	559	2 S33724	transcription fact
42	104.5	3.6	717	2 H85171	DEF (CLA1) protein
43	104	3.6	395	2 S50986	MAF1 protein - yea
44	104	3.6	715	2 H90977	hypothetical prote
45	104	3.6	719	2 F85824	hypothetical prote

ALIGNMENTS

RESULT 1

VCPV3A  
coat protein - adeno-associated virus type 2  
C:Species: adeno-associated virus type 2  
C:Date: 05-Apr-1983 #sequence\_revision 05-Apr-1983 #text\_change 16-Jul-1999  
C:Accession: A03698  
R: Srivastava, A.; Lusby, E.W.; Berns, K.I.  
J. Virol. 45, 555-564, 1983  
A:Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.  
A:Reference number: A03694; MUID:83164299; PMID:6300419  
A:Accession: A03698  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-504 <SRI>  
A:Cross-references: EMBL:J01901; NID:G209616; PIDN:AAA42376.1; PID:G209621; EMBL:M12405;  
C:Superfamily: adeno-associated virus coat protein  
C:Keywords: coat protein

Query Match 74.9%; Score 2177.5; DB 1; Length 504;  
Best Local Similarity 82.5%; Pred. No. 2.1e-153;  
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;

Qy	1	MASGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVITTSRTWALPTYNHLYKQISS	60
Db	1	MATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVITTSRTWALPTYNHLYKQISS	60
Qy	61	ASTGASNDNHFGYSTPWGYPDFNRFCHFSRDMQRLINNNWGFPRKRLNFKLFNIQVK	120
Db	61	QS-GASNDNHFGYSTPWGYPDFNRFCHFSRDMQRLINNNWGFPRKRLNFKLFNIQVK	119
Qy	121	EVTNDGVTIANNLTSTVQVFSDEYQLPVYLGSAHQCLPPPADVFMIPQGYLTLN	180
Db	120	EVTQNDGVTIANNLTSTVQVFSDEYQLPVYLGSAHQCLPPPADVFMIPQGYLTLN	179
Qy	181	NGSQAVERSSFCLEYFPSSOMLRTGNNTFTSYTPEEVPFHSYAHSSOSLDRLMPLIDQY	240
Db	180	NGSQAVERSSFCLEYFPSSOMLRTGNNTFTSYTPEEVPFHSYAHSSOSLDRLMPLIDQY	239
Qy	241	LYYLNRTQNGSGAQNKDLLFSGRSPAGMSVQPKWMLPGPCYQORVSKTKTDNNNSFT	300
Db	240	LYYLSRTNTFTSGTTTQSRLOFSQAGASDIRDSQSNWLPFGPCYQORVSKTSADNNNSYS	299
Qy	301	WTGASKYNLNGRESIINPGTAMASHKDEDKFPPMSGVMI FGKESAGASNTALONVMTD	360
Db	300	WTGATKHLNGRDSLVP--AMASHKDEEKFPPQSGVLIFGKGSEKTNVIERKVMITD	357
Qy	361	EEBIKATNPATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGPWAKI	420
Db	358	EEEGTNPVATEQYGVSVNLQRNQAAATADVNTQGVLPGMWQDRDVLQGPWAKI	417
Qy	421	PHTGCHFPSPMLGGFGIKNPPPOILLIKNTVPVNPAPAEFSATKPSFITQYSTG	475

Db 418 PHTDGHFSPMLGGFGLKHPPIKNTVPANPSTTTFSAAKFPASFTQYSTG 472

## RESULT 2

S52210  
coat protein VP1 - muscovy duck parvovirus  
N:Alternate names: VP1 protein  
C:Species: muscovy duck parvovirus  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-2000  
C:Accession: S52210  
R:Zadori, Z.; Erdel, J.; Nagy, J.; Kisari, J.  
submitted to the EMBL Data Library, September 1993  
A:Reference number: S52209  
A:Accession: S52210  
A:Molecule type: DNA  
A:Residues: 1-732 <ZAD>  
A:Cross-references: EMBL:X75093; NID:G609091; PIDN:CAAS2984.1; PID:G609093  
A:Experimental source: strain FM  
C:Genetics:  
A:Gene: VP1  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein

Query Match 57.7%; Score 1676; DB 2; Length 732;  
Best Local Similarity 56.9%; Pred. No. 4.9e-116;  
Matches 308; Conservative 77; Mismatches 142; Indels 14; Gaps 6;  
QY 1 MASGGAPMADNNEGADGVGNAGNHCDSWLGDRVITTSRTWALPTNNHLYKQISS 60  
DB 199 MAEGSGAMGDSAGGADGVGNAGNHCDSWLGDRVITTSRTWALPTNNHLYKQISS 258  
QY 61 ASTGASNDNHYFGYSTPMGYDFNRFCHFSPRDQRLNNWGRPKRLAFKLFNIQVK 120  
DB 259 GTNPDSN-TQYAGYSTPMGYDFNRFCHFSPRDQRLNNWGRPKRLAFKLFNIQVK 317  
QY 121 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQGLPPFPADVFMIPOYGLTLN 180  
DB 318 EVTTDQTKTIANNLTSTQIFDNEHQLPYVLGSAHQGLPPFPADVFMIPOYGLTWH 377  
QY 181 --NGSOAVGRSSFYCLEYFPQMLRTGNFTFSYTFEVPFHSSYAHQSQSLRLNPLI 237  
DB 378 TWQSGARFNDRAFYLEYFPQMLRTGNFTFSYTFEVPFHSSYAHQSQSLRLNPLI 437  
QY 238 DQYLYLNRQNSGAQNKDLFSGRSPAGMSVQPKWLPQCYRQORVSKTK--TDNN 295  
DB 438 DQYLMNFSEV-NGGRNAQ-----FKKAVKGAFCAGMRNWLPGKLLDQVRAYSQGTNY 491  
QY 296 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSGVMIFGKE--SAGASNTAL 353  
DB 492 ANMSINSGKNKVFLLKDREYLLQPGPVATHTTSDQASSVPAQNIIGIAKDPYRSGSTLAGI 551  
QY 354 DNVMITDEEIKATNPVATERGTVAVNFQSSSDPATGDVHAMGALPGMWQDRDVLQ 413  
DB 552 SDIMVTDEQEIAPTNGVGRHPYGLTVTNEQNTTAPTNAELVGLNGLPGMWQDRDVLQ 611  
QY 414 GPIWAKIPHTDGHFSPMLGGFGLKHPPIKNTVPANPSTTTFSAAKFPASFTQYSTG 473  
DB 612 GPIWAKIPHTDGHFSPMLGGFGLKHPPIKNTVPANPSTTTFSAAKFPASFTQYSTG 671  
QY 474 TQGVSEIEWELQKSKENWNEVQYTSNYAKSANVDFTVDNGLYTEPRPITGTYLTP 533  
DB 672 TQGVSEIEWELQKSKENWNEVQYTSNYAKSANVDFTVDNGLYTEPRPITGTYLTP 731  
QY 534 L 534  
DB 732 L 732

## RESULT 3

VCPVBS  
coat protein VP1 - bovine parvovirus  
N:Contains: coat protein VP2  
C:Species: bovine parvovirus

C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
C:Accession: A26104  
R:Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.  
J. Virol. 60, 1085-1097, 1986  
A:Title: Complete nucleotide sequence and genome organization of bovine parvovirus.  
A:Reference number: A26104; MUID:87061184; PMID:3783814  
A:Accession: A26104  
A:Molecule type: DNA  
A:Residues: 1-673 <CHE>  
A:Cross-references: EMBL:M14363; NID:G333454; PIDN:AAB59847.1; PID:G808805  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F:138-673/Product: coat protein VP2 #status predicted <VP2>

Query Match 17.8%; Score 517; DB 1; Length 673;  
Best Local Similarity 25.1%; Pred. No. 2.9e-30;  
Matches 143; Conservative 85; Mismatches 208; Indels 134; Gaps 17;  
QY 5 GGAPMADNNEGADGVGNAGNHCDSWLGDRVITTSRTWALPTNNHLYKQISSASTG 64  
DB 163 GSGSVGGGGRGGGSGYGTGGTGTTFSENIVVTKNTROPICDKNGHLYKS-EVLNTG 221  
QY 65 ASNDNHYFGYSTPMGYDFNRFCHFSPRDQRLNNWGRPKRLAFKLFNIQVKVTT 124  
DB 222 DTAHQY-AITTPMSYFNFQYSSHSFNDWQHLVNDYERPRPKAMLRVNTLQIKQIMT 280  
QY 125 NDGVTTIANN-LTSTVQVSDSEYQLPYVLGSAHQGLPPFPADVFMIPOYGLTL- 179  
DB 281 DGAMGTYNNDLTAGMHIFCDGDRHYFYVQHPWDQCMPELPSNIWELPQYAYIPAPISV 340  
QY 180 --NNGSOAVGRSSFYCLEYFPQMLRTGNFTFSYTFEVPFHSSYAHQSQSLRL 232  
DB 341 VDNTTNTVBEHLKGVPLYLENSDHEVLRNG-----AGMSYQPKNWL 375  
QY 233 MNPLIDQYLYLNRQNSGAQNKDLFSGRSP-----SAGASNTALDNVMITDEEIKATNPVAT 277  
DB 376 YRIYQLWRUMRDKRHHQHASDDVQSTQCKQKLLIQRKQNKORPQAAUATSNWM 435  
QY 278 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSG 337  
DB 436 SGP-----GIARGHTNATLQTSAGALVTWVTNGAD----VSG 469  
QY 338 VM-----IFGKE-----SAGASNTALDNVMITDEEIKATNPVAT 372  
DB 470 VRAVRGVSTDPYGGQSPESDRLRLRYSAASAEQQNPILN-----AAR 515  
QY 373 ERFGTVAVNFQSSSDPATGDVHAMGALPGMWQDRDVLQGPITWAKIPHTDGHFSP 432  
DB 516 HTFTREARTKLITGSGADGDYKEWMLPNQWMSAPISRYNPIWVPRVNRKTLDDTQ 575  
QY 433 MGFGGLKNPPPIQILIK--NTPVPANPSTTTFSAAKFPASFTQYSTGVSVEIWELEQKENS 490  
DB 576 DGSIPMSHPPTIFIKLARIPTVPCNGD-----SFLNIYVTGQVSCVYVWEVEKRG 626  
QY 491 KRWNPVQYTSNYAKSANVD-FTVDNGLY 519  
DB 627 KRWNPVQYTSNYAKSANVD-FTVDNGLY 653

## RESULT 4

VCPV19  
coat protein VP1 - parvovirus B19 (strain Au)  
C:Species: parvovirus B19  
A:Note: host Homo sapiens (man)  
C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
C:Accession: A24299  
R:Shade, R.O.; Blundell, M.C.; Cotmore, S.F.; Tattersall, P.; Astell, C.R.  
J. Virol. 58, 921-936, 1986  
A:Title: Nucleotide sequence and genome organization of human parvovirus B19 isolated from  
A:Reference number: A24299; MUID:86200451; PMID:3701931  
A:Accession: A24299  
A:Molecule type: DNA  
A:Residues: 1-781 <SHA>







Db 263 EYATGYTFDEIRLTHWTNRHLQPPQITELPSSDTANATLTARGYRSLGTQIOGR 322  
QY 250 -----QSGSAQND-LLFSGSGPAGMSVQPKWLPG-----PCYRQOR 286  
Db 323 NQVTEATRVPAQVGCQPHDNFETGRAGFKVPVVPADITQGLDHDANGSLRYTYDKOH 382  
QY 287 VSKTKTDNNNSNFTWTGASKYINLNGRESIIPGCTAMASHDKDDKFFPMGVMIFKESA 346  
Db 383 GQSWASQNNKDRYTM-DAVNYD-SGR----- 406  
QY 347 GASNTALDNVMTDEBEIKATNPVATER---FGTVAVNFOS---SSTDPAATGDVHAMGALP 401  
Db 407 WTNNCPIQSPVPTSEPN---ANOILNRDNLAGKTDIHTNAFNSYGLTAPFPH----- 457  
QY 402 GMVWQDRDVLQGPPIWAK---IPHTDGHFHPSPLMGGFGLKNPPQILLIKNTP--VPANP 456  
Db 458 -----PAPIYPOQIWDKELDLHKPRLHTQAPFV---CKNNAPGQLVRLAPNLTDQXD 509  
QY 457 PAEFSATKPAFTQYSTQGVSEIWEIWELOKENSKRWNPEVOYTS 501  
Db 510 PNSSNLSRIYVTGTFFWKGLTLKAK---MRPNA-TWNPVFOISA 550

RESULT 9  
VCPVW2  
coat protein VP1 - parvovirus H1  
C:Species: parvovirus H1  
A:Note: host Homo sapiens (man)  
C:Date: 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change 08-Apr-1994  
C:Accession: A03699  
R:Rhode III, S.L.; Paradiso, P.R.  
J. Virol. 45, 173-184, 1983  
A:Title: Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybridization  
A:Reference number: A03695; MUID:83112183; PMID:6823009  
A:Accession: A03699  
A:Molecule type: DNA  
A:Residues: 1-722 <RHO>  
A:Cross-references: EMBL:X01457; EMBL:J02198  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein

Query Match 7.8%; Score 226; DB 1; Length 722;  
Best Local Similarity 21.7%; Pred. No. 1.1e-08;  
Matches 128; Conservative 83; Mismatches 234; Indels 146; Gaps 26;  
QY 2 ASGGAPMADNNEGADGVGNASGNWHCDSTW-LGDRVITTSRTWAL----- 47  
Db 153 ADGGGS-SGGGGGGGGIGVSTGYDQNTYTKFLGQVWEITAHASRLHLHGMPPSENYC 211  
QY 48 --PTYNHLYKQISSASTGASNDNHFGYSTPMGYDFNRFCHFSPRDWQRLNNMGF 105  
Db 212 RVTVHNNQTTGHTKVGKNMAYDTHQIWI-TWMSLVANANGVWFQPSDWQFIQNSMESL 270  
QY 106 RKELNFKLNTQVKEVTT-----NDGVTIANLSTVQVSDSEYQLPYVLGSAHQOC 160  
Db 271 NUDLSQELFNVVVKTVTSCQAGAGDAIKVYNDLTACHMMVALDSNNILPYTPAAQTSST 330  
QY 161 L-----PPPAD---VFMPQVGLTLNNGSQ-----AVG-----RSSFCLE-YPFS 199  
Db 331 LGFYPKWPTAPARYFYFPMRQLSVTSNSAEGTQITIGEPQALNSQFTIENTLFI 390  
QY 200 QMLRTGNFTF-SYTFEEVFFHSSYA-----HSQSLDRLMN-PLIDQYLYLNRTQNSQSG 252  
Db 391 TLLRTGDEPTGTYIPNTDPLKLTHTWQNRHLACLQGITDLTSDTATASLTANGDRFG 450  
QY 253 SAQNKDLP-----SRSPAGMSVQPKNWLPGPCYRQORVSK 289  
Db 451 STQTONVNTVTEALRTPAQIGFMQPHDNFANRGGPFKVPVVP-----LDI 497  
QY 290 TKTDDNNSNFTWTGASKYINLNGR---ESIINPGTA-----MASHKDDKDEKFFPMG 337  
Db 498 TAGEDHDAN---GAIRFYQKHQGEDWAKQGAAPERYTWDAIDSAAGRDRTARCFV----- 549

QY 338 VMIFGKESAGSANTALDNVMTDEBEIKA-TNPVATERFGTVAVNFQSSSTDPATGDVHA 396  
Db 550 -----QSAPISIPNQIQLQREDAIAGRTHMTNTVFNSTVNSYGLSAPPHDP----- 596  
QY 397 MGALFPMWQDRDVLQGPPIWAK---IPHTDGHFHPSPLMGGFGLKNPPQILLIKNTPVP 453  
Db 597 -----IYPNGQIWDKELDLHKPRLHTVAPPFV---CKNNPPGQLFVHLGP-- 638  
QY 454 ANPPAEF--SATKPAFTQYSTQGVSEIWEIWELOKENSKRWNPEVOYTSN 502  
Db 639 -NLTDQFDNPNSTVSRIVT-YSTFYWKGLKFKAKLRPNLTWNPVYQATTD 687

RESULT 10  
VCPV2M  
coat protein VP1 - minute virus of mice  
C:Species: minute virus of mice, murine parvovirus  
C:Date: 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change 08-Apr-1994  
C:Accession: A03700  
R:Astell, C.R.; Thomson, M.; Merchlinsky, M.; Ward, D.C.  
Nucleic Acids Res. 11, 999-1018, 1983  
A:Title: The complete DNA sequence of minute virus of mice, an autonomous parvovirus.  
A:Reference number: A03696; MUID:83143341; PMID:6298737  
A:Accession: A03700  
A:Molecule type: DNA  
A:Residues: 1-716 <AST>  
A:Cross-references: EMBL:V01115  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein

Query Match 6.8%; Score 197; DB 1; Length 716;  
Best Local Similarity 20.0%; Pred. No. 1.6e-06;  
Matches 126; Conservative 88; Mismatches 241; Indels 174; Gaps 28;  
QY 3 SGGAPMADNNEGADGVGNASGNWHCDSTW-LGDRVITTSRTWALPYNNHLYKQ 57  
Db 160 SGGG-----GSGGGGVGVTGSDYDQNTYTKFLGQVWEITAHASRLHLHGMPPSENYC 213  
QY 58 I-----SSASTGASNDNHFGYSTPMGYDFNRFCHFSPRDWQRLNNMGFPRPKEL 110  
Db 214 IRVHTTDTSVKGNMAKODAHQIWTWMSLVANANGVWLQPSDWQYICNTWSQLNLSL 273  
QY 111 NPKLFNIQVKEVTTND---GVTIANLSTVQVSDSEYQLPYVLGSAHQCLPPPPA 166  
Db 274 DQEIFNVVLKVTQEDLGGQAIKIYNDLTACHMMVALDSNNILPYTPAANSMETLGFYPM 333  
QY 167 DVFMTPQVGY-----LTLNNGSQV-----GRSFCYCLEYFPPS-QMLRTGN 207  
Db 334 KPTIASPYRYPCVDRDLSVTYENQEGTVENVMGTGKIGIQQFTIENTQITLLRTGDE 393  
QY 208 F-TFSYTFEEVFFHSSYAHSQSLDRLMN-PLIDQYLYLNRTQNSGSAQNKOLLFSRG 264  
Db 394 FATGYTFDTHSV--KLTHWTQNRHLQCPPLLSTF---PEADTDAGT-----LTAQG 441  
QY 265 SPAGMSVQPKNWL-----EGPCYRQORVSKTKTDNNNSNFT 300  
Db 442 SRHGTQMGVNVWVSBARTPAQVGCQPHDNFASRAGP-PAAPKVPADITQGVDEKEAN 500  
QY 301 WTGASKYINLNGRESIINPG-----TAMASHKDDKDEKFF-----PMGVMIF 341  
Db 501 GSVRYSYQKHQENWASHGAPERYTWDETSPGSRDTKDGIQSAPLVVPPLNGI--- 557  
QY 342 GKESAGSANTALDNVMTDEBEIKATNPVATERFGTVAVNFOS---SSTDPAATGDVHAMGA 399  
Db 558 -----LTNANPITGN-----DIHFSNVNSVGLTAPFPH----- 587  
QY 400 LFGMWQDRDVLQGPPIWAK---IPHTDGHFHPSPLMGGFGLKNPPQILLIKNTP--VPA 454  
Db 588 -----PSPVYPOQIWDKELDLHKPRLHTVAPPFV---CKNNAPGQLVRLGNLTDQ 637  
QY 455 NPPAEFSATKPAFTQYSTQGVSEIWEIWELOKENSKRWNPEVOYTSYKASNVDTFTD 514  
Db 638 YDPNGATLSRIYVTGTFFWKGLTLKAKRA-----NTTNVYQVSAE-----D 682

QY 515 NNGLY---TEPRPIGT-----RVLTREPL 534  
Db 683 NGNSYMSVTKMLPTATGNMQSVPLITRPV 711  
RESULT 11  
VCPVIM  
coat protein VP1 - minute virus of mice (strain VVM1)  
C:Species: minute virus of mice, murine parvovirus  
C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 28-Jul-2000  
C:Accession: B23008; B29510  
R:Salhi, R.; McMaster, G.K.; Hirt, B.  
Nucleic Acids Res. 13, 3617-3633, 1985  
A:Title: DNA sequence comparison between two tissue-specific variants of the autonomous  
A:Reference number: A23008; MUID:85242059; PMID:3855242  
A:Accession: B23008  
A:Molecule type: DNA  
A:Residues: 1-718 <SAH>  
A:CROSS-references: EMBL:X02481; NID:G60918; PIDN:CAB46507.1; PID:G5419928  
R:Atwell, C.R.; Gardiner, E.M.; Tattersall, P.  
J. Virol. 57, 656-669, 1986  
A:Title: DNA sequence of the lymphotropic variant of minute virus of mice, VVM(ii), and C  
A:Reference number: A29510; MUID:86115415; PMID:3502703  
A:Accession: B29510  
A:Molecule type: DNA  
A:Residues: 1-143, 'A', 145-718 <AST>  
A:CROSS-references: EMBL:M12032  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein

Query Match 6.8%; Score 196.5; DB 1; Length 718;  
Best Local Similarity 20.5%; Pred. No. 1.7e-06;  
Matches 122; Conservative 82; Mismatches 243; Indels 147; Gaps 25;  
QY 3 SCGGAPMADNNEGADGVGNASGNHCHDS--TWLGD---RVITSTRTWALPTYNHLYKQ 57  
Db 161 SGGG-----GGGGGCVSTGSGSYNQTHYRFLGPGWVEITATRLVHLNMPKSENCR 214  
QY 58 I-----SSASTGASNDNHFGYSTPGYFDFNPFCHFSPRDWRQRLNNWGPRL 110  
Db 215 IRVHTTDTSVKGNWAKDHAHQIWTPLSLVDANAWGLQSDUNQYICNTMSQLNLVSL 274  
QY 111 NPKLFNIQVETVND---GVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGLPPPPA 166  
Db 275 DQEIFNVVLKTVTEQDSGGQAIKIYNNDLTACMVAVDNNILPTPAANSMETLGFYFW 334  
QY 167 DVPMIPQY-----LTNN-----GSQAVGRSSFYCLEYFPPS-QMLRTGN 206  
Db 335 KPTIASPYRYFCVDRDLSTVYENQEGTIEHNVMTGPKGMNSQFTIENTQOITLLRTGD 394  
QY 207 NF-TFSYTPPEVPFSSVAHSQSRLNN--PLIDQYLYLNRQNSGSAQNKDLLFSR 263  
Db 395 EFATGYTFDTPV--KLTHWTQNRQLGQPLLSTF---PEADTAGT-----LTAQ 442  
QY 264 GSPAGMSVOPKNL-----PGPCYRQORVSKTKTDNNNNNF 299  
Db 443 GSRHGATOMEVNVWSEAIRTPAQVFCQPHNDFEASRAGP-FAAPKVPADVTQGVDRFA 501  
QY 300 TWTGAKYNLNGRESIINPGTAWASHKDEDF-----FPMGVMIFGKESAGASN 350  
Db 502 NGSVRSYKQNGENWAAAGPAPERTWDTNFGGRDRTRDGFIOAPLVVPPPLNGILT 561  
QY 351 TALDNVMTDEBEIKATNPVATERFGTAVNFQS--SSTDPAATGDVHAMGALPGVMWQDR 408  
Db 562 NA-----NPIGKN---DIHFSNVNSYGLPTAFSH-----PS 591  
QY 409 DVLQGPWIAK---IPTHGCHFPSPMLGFGFLGNPPQIILKNTP--VPANPPAEFSAT 463  
Db 592 PVPYQCGIWDKELDLHFKRLHITAPFV---CKNNAPGQMLVRLGNLTDQDPPNGATLS 648  
QY 464 KPAFTITQYSTGVSEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 517

Db 649 RIVTYGTFFWKGKLTWRAKLRA-----NTTNVPYQ-----VSVEDNG 686  
RESULT 12  
VCPVME  
coat protein VP1 - mink enteritis virus (strain Abashiri)  
N:Contains: coat protein VP2  
C:Species: mink enteritis virus, MEV  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 12-Apr-1996  
C:Accession: B38350  
R:Kariatsumari, T.; Horiuchi, M.; Hama, E.; Yaguchi, K.; Goto, H.; Shinaga  
J. Gen. Virol. 72, 867-875, 1991  
A:Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the  
A:Reference number: A38350; MUID:91202123; PMID:2016597  
A:Accession: B38350  
A:Molecule type: DNA  
A:Residues: 1-722 <KAR>  
A:CROSS-references: GB:D00765  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F:139-722/Product: coat protein VP2 #status predicted <VP2>  
Query Match 6.7%; Score 194; DB 1; Length 722;  
Best Local Similarity 20.7%; Pred. No. 2.7e-06;  
Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;  
QY 5 GGAPMADNNEGADGVGNASGNHCHDS--TWLGD---RVITSTRTWALPTYNHLYKQ 52  
Db 148 GGQP-AVNERATGSGNGSGGGGGG--GGVIGISTGTFFNNQTEFKLENGWVEITANSS 204  
QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPGYFDFNPFCHFSPRDW 95  
Db 205 RLVLHNPSESNYKRVVNVNNDKTAVKGNMALDDTHVQIVTPMSLVDANAWGVNFPGDW 264  
QY 96 QRLNNWGPRLNFKLNIQVKEVT---TNDGVTTIANNLTSTVQVFSSEYQLPYV 152  
Db 265 QLVNTWSEHLVSEFQEIFNVVLKTVSESATQPTKYVNDLTASLWALDLSNNTWPT 324  
QY 153 LGSAAHQCL-----PPFPA-----DVPMIPQYGYLTLNNGSAQVGRS- 189  
Db 325 PAAMRSETLGFYPMKPTIPTPRYFQWDRTLIPSH-----TGTSGTPTNIYHGTDPDD 378  
QY 190 -SFYCLE-YPPSOMLRGNP---TF-----SVTFEE-----VPHSSSYAHSQSL 229  
Db 379 VQFYTIENSVPVHLRTGDEFATGTFFDCPKCLTHWTNRALGLPPLNLSLPQSEGA 438  
QY 230 DR-----LWNPIDQYL--YYLNRQNS----- 251  
Db 439 TNFGDIGVQODKRGVTQMGNTDYITEATIMRPAEYGSAPYYSFEASTQGPPTIAAG 498  
QY 252 -GSAQNKDLLFSRGSAGMSVQPKMWLPGPY---RQORVSKTKTDNNNSNFTWTGASKY 307  
Db 499 RGGATDENOAAADGPR-----YAFGRHGQKQTTTTTGETPFTY----- 538  
QY 308 NLNGRESIINPGTAWASHKODEDKFFPMGVMIFGKESAGASNALD-NVMTDDEEIK 366  
Db 539 -----IAHQDT-----GRYPAGDWIIONINFLNFTVNDVLLP 570  
QY 367 TNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGVMWQDRVYLOQPIWAKIPTHDTGH 426  
Db 571 TDPIG---GKTGINY--TNIFNTYGLPTALNNVP-----PVYNGQIWDKPFDTD-- 615  
QY 427 FHPSPLMGGFGLK-----NPPQIILKNTPVPA-----PPAEFSATPEASIT 470  
Db 616 -----LKRLHVNPAFVQNNCPGLFVKVAPNLNTNEYPDASAMSR-----IV 660  
QY 471 QYSTGVSEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIG 526  
Db 661 TYSDFWKGKLVFKAKLRASHTWNPQIOMSIN-----VDNQFNYL--PNNIG 705  
RESULT 13  
VCPVLP

coat protein VP1 - feline panleukopenia virus  
 N:Contains: coat protein VP2  
 C:Species: feline panleukopenia virus, FPLV  
 C>Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
 C/Accession: A03701  
 R:Carlson, J.; Rushlow, K.; Maxwell, I.; Maxwell, F.; Winston, S.; Hahn, W.  
 J. Virol. 55, 574-587, 1985  
 A:Title: Cloning and sequence of DNA encoding structural proteins of the autonomous parvovirus  
 A:Reference number: A03697; MUID:85265017; PMID:2991581  
 A:Accession: A03701  
 A:Molecule type: DNA  
 A:Residues: 1-727 <CR>  
 A:CROSS-references: EMBL:M10824; NID:G333474; PIDN:AAA47161.1; PID:G333476  
 C/Genetics:  
 A:Introns: 11/1  
 C:Superfamily: parvovirus coat protein  
 C:Keywords: coat protein  
 P:144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.7%; Score 194; DB 1; Length 727;  
 Best Local Similarity 20.7%; Pred. No. 2.7e-06;  
 Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;

QY 5 GGAPMADNNEGADGVGNAGNHCDSWLGDRVITST-----RTWALPTVNN- 52  
 DB 153 GGQP-AVRNERATGSGNGSGGGGGGS--GGVGISTGTFTNNQTEPKFLENGWVEITANS 209  
 QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPMGYDFDNRFCHFSRDPW 95  
 DB 210 RLVLNMPSENYRVRVNNMDKTAVGNMALDDTHAQIVTPNSLVDANAGWVFNPGDW 269  
 QY 96 QRLINNMGFRPKRLNFKLFNIQVKEVT---TNDGVTTIANNTSTVQVFSSEYQLPYV 152  
 DB 270 QLIVNTMSLHVSFEQEIFNVVLKTVSESATQPTKVYNNDLTASLWALDSNNTMPT 329  
 QY 153 LGSAGQCL-----PPFPA-----DVFMIPQGYLTLLNGSOAVRS----- 189  
 DB 330 PAAMRSETLGFYPMKPTIPTWRYFQWDRTLPSH-----TGTSGTPTNIYHGTDPPD 383  
 QY 190 -SFYCLE-YFPQMLRTGNF---TF-----SYTPEE-----VPHSSYAHQSOL 229  
 DB 384 VQFYTIENSVPVHLRTGDEFATGTFDFCKPCLRLTHWTQNRALGLPPLNSLPQSEGA 443  
 QY 230 DR-----LMNPLIDQYL---YYLNRQNGS----- 251  
 DB 444 TNFGDIGVQDKRRGVQMGNTDYITEATIMRPAEYGSAPYYSFEASTGPKPIAAG 503  
 QY 252 -GSAQNKDLLPSRGSAGMSVQPKWLPGPCY---RQORVSKTKTNNNSNFTWTGASKY 307  
 DB 504 RGAQTDENQADGDR-----YAFGRHQGQKTTTGTPTERTY----- 543  
 QY 308 NLNGRESIIPGTAMASHKHDEKDFPMSGVMIFGKESAGASNTALD-NVMTIDBEEIK 366  
 DB 544 -----IAHQDT-----GRYPAGDWIQNINFLPVTDNVLPL 575  
 QY 367 TNPVATERFTVAVNFOSSTDPATGDVHANGALPGMWQDRDVLQGPPIWAKIPHTDGH 426  
 DB 576 TDPIG-----GKTGINY--TNIFTYGLTALNNVP-----PVYPNGQIWDKEFTD-- 620  
 QY 427 FHPSPLMGGFLK-----NPPQILIKNTVPAN---PPAEFSATKPFASFIT 470  
 DB 621 -----LKPRHLVNAFPVQNNCPGQLFVKVAPNLTYEYDSDASANMSR-----IV 665  
 QY 471 QYSTGQVSVIEWELQKENSKRWNPEVQYTSNYSKASANVDFTVDNNGLYTEPRPIG 526  
 DB 666 TYSDFWFKGLVFKAKLRASHTNWPIQMSIN-----VDNQENYL-PNNIG 710

RESULT 14  
 S49594  
 capsid protein VP2 - canine parvovirus  
 C:Species: canine parvovirus, CPV  
 C>Date: 05-Mar-1995 #sequence\_revision 10-Nov-1995 #text\_change 26-Aug-1999

C/Accession: S49594  
 R:Mizak, B.; Plucienniczak, A.  
 submitted to the EMBL Data Library, November 1994  
 A:Description: Polish isolates of canine parvovirus.  
 A:Reference number: S49594  
 A:Accession: S49594  
 A:Molecule type: DNA  
 A:Residues: 1-584 <MIZ>  
 A:CROSS-references: EMBL:Z46651; NID:G572660; PIDN:CAA86612.1; PID:G572661  
 C:Superfamily: parvovirus coat protein

Query Match 6.5%; Score 189; DB 2; Length 584;  
 Best Local Similarity 20.8%; Pred. No. 4.6e-06;  
 Matches 132; Conservative 89; Mismatches 223; Indels 192; Gaps 33;

QY 5 GGAPMADNNEGADGVGNAGNHCDSWLGDRVITST-----RTWALPTVNN- 52  
 DB 10 GGQP-AVRNERATGSGNGSGGGGGGS--GGVGISTGTFTNNQTEPKFLENGWVEITANS 66  
 QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPMGYDFDNRFCHFSRDPW 95  
 DB 67 RLVLNMPSENYRVRVNNMDKTAVGNMALDDTHAQIVTPNSLVDANAGWVFNPGDW 126  
 QY 96 QRLINNMGFRPKRLNFKLFNIQVKEVT---TNDGVTTIANNTSTVQVFSSEYQLPYV 152  
 DB 127 QLIVNTMSLHVSFEQEIFNVVLKTVSESATQPTKVYNNDLTASLWALDSNNTMPT 186  
 QY 153 LGSAGQCL-----PPFPA-----DVFMIPQGYLTLLNGSOAVRS----- 189  
 DB 187 PAAMRSETLGFYPMKPTIPTWRYFQWDRTLPSH-----TGTSGTPTNIYHGTDPPD 240  
 QY 190 -SFYCLE-YFPQMLRTGNF---TFSTVTEEVPHSSYAHQSOLRLMNLIDQILYLYLNR 246  
 DB 241 VQFYTIENSVPVHLRTGDEFATGTFDFCKP--CRLTHWTQNRALG--LPPPLNSLPQ 296  
 QY 247 TQNS-----GSAQNK-----DLLFRSGSPAGMSVQPKWLPGPCVQRQRY 287  
 DB 297 SEGTFNGYIGVQDKRRGVQMGNTDYITEATIMRPAEYGS-----APYYSFEAS 348  
 QY 288 SK-----TKTDNNNSNFTWTGASKY---NLNGRESIIPGT---AMASHKD 327  
 DB 349 TGQPKPTPIAAGRGAQTDENQAA---GDPRYAFGRHQGQKTTTGTPTERTYIAHOD 405  
 QY 328 DEDKFFPMSGVMIFGKESAGASNTALD-NVMTIDBEEIKATNPVATERFTVAVNFQSS 386  
 DB 406 T-----GRYPGDIQININFLPVTDNVLPLTDPIG---GKTGINY--TN 446  
 QY 387 TDPATGDVHANGALPGMWQDRDVLQGPPIWAKIPHTDGHFHPSPLMGGFLK----- 439  
 DB 447 IFNTYGLTALNNVP-----PVYPNGQIWDKEFTD-----LKPRHLVNA 486  
 QY 440 -----NPPQILIKNTVPAN---PPAEFSATKPFASFITQYSTGQVSVIEWELQKENS 490  
 DB 487 PFVQNNCPGQLFVKVAPNLTYEYDSDASANMSR-----IVTYSDFWFKGLVFKAKLRAS 542  
 QY 491 KRWNPVQYTSNYSKASANVDFTVDNNGLYTEPRPIG 526  
 DB 543 HTWNPQMSIN-----VDNQENYL-PSNIG 567

RESULT 15  
 VCPVFP  
 coat protein VP1 - feline panleukopenia virus (strain 193)  
 N:Contains: coat protein VP2  
 C:Species: feline panleukopenia virus, FPLV  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 12-Apr-1996  
 C/Accession: B36608  
 R:Martyn, J.C.; Davidson, B.E.; Studdert, M.J.  
 J. Gen. Virol. 71, 2747-2753, 1990  
 A:Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvovirus  
 A:Reference number: A36608; MUID:91073139; PMID:2174965  
 A:Accession: B36608  
 A:Molecule type: DNA

A;Residues: 1-727 <MAR>  
A;Cross-references: GB:X55115  
C;Superfamily: parvovirus coat protein  
C;Keywords: coat protein  
F;144-727/Product: coat protein VP2 #status predicted <VP2>

	Query Match	6.5%;	Score 189;	DB 1;	Length 727;
	Best Local Similarity	20.6%;	Pred. No. 6.3e-06;		
	Matches 135;	Conservative 73;	Mismatches 216;	Indels	232; Gaps 32;
Qy	5	GGAPMADNNEGADGVGNASGNHCHDS	TWLGDRVTTST-----	RTWALPTVNN-	52
Db	153	GGQP-AVNERATSGSGSGGGGGGS-	GGVGISTGTFFNQTEPKFLENGVEITANS		209
Qy	53	---HL-----YKQI-----	SSASTGASNDNHFGYSTPMGVDFPNRPHCFSPRDW		95
Db	210	RLVHLNMPSESEYKRVVNNNDKTA	VKGNMALDDIHVQIVTPWSIVADANAGVWFNPGDW		269
Qy	96	QRLINNNGWFPKRLNFKLFINIQKEVT--	TNDGVTTIANNLSTVQVFSDESQLPYV		152
Db	270	QLIVNTWSELHVSFEQEIFNVLKTVSESATQ	PKFYNNNDLTASLVALDSSNNTMPTT		329
Qy	153	LGSAAHOGCL-----PPPPA-----	DVFMIPOGYLTLNNGSQAVGRS-----		189
Db	330	PAAMRSETLGFYPKPKTPTTPWRYFQWDR	LILPSH-----TGTSCTPTVNVYHCTDPDD		383
Qy	190	-SPYCLE-YFPSQMLRTGNPF--TF-----	SYTFEE-----VFFHSSYAHSQSL		229
Db	384	VQFYTIENSVPVHLRLTGDEFATGTFP	FDCKPCRLTHWTQNRALGLPFLNSLPQSEGA		443
Qy	230	DR-----LNNPLDQYL--YLLNRTQNQS	-----		251
Db	444	TNFGDIGVQDKRGVGTQMGNTDYIT	EATIMKPAEYSGVSAPIYSFEASTQGFPTPIAG		503
Qy	252	-GSAQNDLLFSRGSFAGMSVQPKNWL	GPCPY---RQORVSKTKTDNNNSNFTWTGASKY		307
Db	504	RGGAQTDENQAADGPR-----	YAFGRHQKQKTTTGETPERFTY-----		543
Qy	308	NLNGRESIINPGTAMASHKHDDK	FPMSGVMIPOKESAGASNTALD-NVMITDSEETKA		366
Db	544	-----IAHQDT-----	GRYPEGDWIQINFINFLPYTNDNVLLP		575
Qy	367	TNPVATERFGTVAVNFQSSSTD	PATGDVHAMGALPGWQDRDVLQGPIMAKI	PHTDGH	426
Db	576	TDPIG-----GKTGINY--TNINFTY	GPLTALANVP-----PVTNGOIWKEP	DTD--	620
Qy	427	FHPSPLMGFGFLK-----	NPPQILIKNTPVPAN---PPAEFSATK	FASTT	470
Db	621	-----LKPRLHVNAFPVCNN	CEGQLFKVKAPNLTWEYDPA	SANMSR---IV	665
Qy	471	QYSTQGVSVIEIWELOKENS	KRWNPDEVQVTSNYAKS	ANVDFTVDNNGLYTEPR	PIG 526
Db	666	TVSDFWWTKGLVFKAKLRASH	TWNPLOOMSIN-----VDNOFN	VV-PNNIG	710

Search completed: January 21, 2004, 16:07:09  
Job time : 14.5714 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 15:59:46 ; Search time 9.14286 Seconds  
(without alignments)  
2746.653 Million cell updates/sec

Title: US-09-807-802A-17  
Perfect score: 2906  
Sequence: 1 MASGGGAPMADNEGADGVG.....NNGLYTEPRIGTRYLTRPL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2177.5	74.9	504	COA3_AAV2	P03135 adeno-assoc
2	517	17.8	673	COAT_PAVBO	P07297 bovine parv
3	487.5	16.8	781	COAT_PAVHB	P07299 human parv
4	232.5	8.0	729	COAL_PAVP9	P33484 porcine par
5	232.5	8.0	729	COAL_PAVPN	P18546 porcine par
6	230.5	7.9	729	COAL_PAVPK	P52501 porcine par
7	229	7.9	587	COAT_PAVL3	P36310 parvovirus
8	226	7.8	722	COAT_PAVHH	P31336 hamster par
9	197	6.8	716	COAT_MUMIV	P03137 murine minu
10	196.5	6.8	718	COAT_MUMIM	P07302 murine minu
11	194	6.7	722	COAT_MEVA	P27437 mink enteri
12	194	6.7	727	COAT_FPV	P04864 feline panl
13	190	6.5	584	COAT_PAVC2	P30129 canine parv
14	189	6.5	727	COAT_FPV19	P24840 feline panl
15	188	6.5	727	COAT_PAVCB	Q11213 canine parv
16	188	6.5	737	COAT_PAVCD	P17455 canine parv
17	184	6.3	748	COAT_PAVCN	P12930 canine parv
18	177.5	6.1	647	COAT_ADVG	P24029 aleutian mi
19	176	6.1	722	COAT_PAVC7	P04863 canine parv
20	122.5	4.2	648	FXN1_HUMAN	O15353 homo sapien
21	121	4.2	880	SVY_BACST	P11931 bacillus st
22	109	3.8	1186	CEAA_BACTS	Q45710 bacillus th
23	108	3.7	655	CDGT_KLEPN	P08704 klebsiella
24	107.5	3.7	648	FXN1_MOUSE	O61575 mus musculu
25	106.5	3.7	1070	Y355_HUMAN	O15063 homo sapien
26	106	3.6	1271	Y338_MTCGE	P47580 mycoplasma
27	105	3.6	1396	IT42_DROME	P12080 drosophila
28	104.5	3.6	559	HNFB_PIG	Q03365 sus scrofa
29	104.5	3.6	717	DXS_ARATH	Q38854 arabidopsis
30	104	3.6	395	MAFI_YEAST	P41910 saccharomyc
31	103.5	3.6	1196	AMVB_PAEPO	P21543 paenibacill
32	102	3.5	1122	ADPI_MTCGA	Q49379 mycoplasma
33	102	3.5	2493	CYAA_UTSMA	P49606 ustilago ma

34	102	3.5	3712	1	LMA_DROME	Q00174 drosophila
35	100.5	3.5	531	1	HEXB_PIG	Q28548 sus scrofa
36	100	3.4	739	1	CATA_MYCSM	Q59557 mycobacteri
37	99	3.4	879	1	GUNI_CLOTM	P23899 rattus norv
38	98.5	3.4	557	1	HNFB_RAT	P40209 saccharomyc
39	97.5	3.4	560	1	GAT2_YEAST	Q66810 ebola virus
40	97.5	3.4	676	1	VGP_EBOIC	Q9V252 homo sapien
41	97.5	3.4	685	1	RNF6_HUMAN	Q02630 saccharomyc
42	97.5	3.4	1113	1	N116_YEAST	Q00955 saccharomyc
43	97.5	3.4	2233	1	COAC_YEAST	Q095741 homo sapien
44	97	3.3	557	1	CNE6_HUMAN	P06564 bacillus sp
45	97	3.3	800	1	GUN_BACSI	

ALIGNMENTS

RESULT 1  
COA3\_AAV2  
ID\_COA3\_AAV2 STANDARD; PRT; 504 AA.  
AC P03135  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1991 (Rel. 20, Last annotation update)  
DE Probable coat protein 3.  
OS Adeno-associated virus 2 (AAV2).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=10804;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83164299; PubMed=6300419;  
RA Srivastava A., Lusby E.W., Berns K.I.;  
RT "Nucleotide sequence and organization of the adeno-associated virus 2 genome.";  
RL J. Virol. 45:555-564 (1983).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: J01901; AAA42376.1; -  
DR PIR: A03698; VCPV3A.  
DR InterPro: IPR001403; Parvo\_coat.  
DR Pfam: PF00740; Parvo\_coat; 1.  
KW Coat protein.  
SQ SEQUENCE 504 AA; 56366 MW; 758998B017052B6B2 CRC64;

Query Match	74.9%	Score	2177.5	DB	1	Length	504
Best Local Similarity	82.5%	Pred. No.	2.4e-149	Indels	3	Gaps	2
Matches	392	Conservative	33	Mismatches	47		
QY	1	MASGGGAPMADNEGADGVGNASGNWCHDSWLGDRVTTSTRTWALPTNNHLYKQLSS	60				
DB	1	MATGSGAPMADNEGADGVGNSSGNWCHDSWLGDRVTTSTRTWALPTNNHLYKQLSS	60				
QY	61	ASTGASNDNHFGYSTPMGYDFNRFCHGSPRQWRQRLNNWGFPRKRLNFKLNIQVK	120				
DB	61	QS-GASNDNHFGYSTPMGYDFNRFCHGSPRQWRQRLNNWGFPRKRLNFKLNIQVK	119				
QY	121	EVTNDGVTTANNLTSTVQVFSDEYOLPVVLSAHOGCLPPPPADVFMLPOVGYLTN	180				
DB	120	EVTQNDGVTTANNLTSTVQVFSDEYOLPVVLSAHOGCLPPPPADVFMLPOVGYLTN	179				
QY	181	NGSQAAGRSSFCLEYFPFQMLRTGNFTFSTYEEVFHSSYAHQSQSLRLMNPDIQY	240				
DB	180	NGSQAAGRSSFCLEYFPFQMLRTGNFTFSTYEEVFHSSYAHQSQSLRLMNPDIQY	239				
QY	241	LYLNRTOQSGSQNKDLLFSRGSAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSPT	300				

Db 240 LYLRTNTPSTTQSLRQFQAGASDIRDQSRNWLPGPCYQORVSKTSADNNNSYS 299  
 QY 301 WTGASKYNLNGRESINPGTAMASHKDDKFFPMGSMVNIKFGESAGASNTALDNVMTD 360  
 Db 300 WTGATKYNLNGRDSLVP--ANASHKDDKFFPMGSMVNIKFGESAGASNTALDNVMTD 357  
 QY 361 EBEIKATNPVATERFGTVAVNFQSSSTDPAQGVHMGALPGVWQDRDVLVQGPWAKI 420  
 Db 358 EBEIGTNPVATEQVSVSTNLQRNRQAADVNTQGVLPQVWQDRDVLVQGPWAKI 417  
 QY 421 PHTDGHFHPSPLMGGFGLKHPQIILKNTVPANPPAFAFSATKASEFTQYSTG 475  
 Db 418 PHTDGHFHPSPLMGGFGLKHPQIILKNTVPANPPAFAFSATKASEFTQYSTG 472

## RESULT 2

COAT\_PAVBO STANDARD; PRT; 673 AA.  
 AC P07297; Q84374;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coat protein VP2 (Contains: Coat protein VP3).  
 OS Bovine parvovirus (BPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=10784;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87061184; PubMed=3783814;  
 RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,  
 RA Bates R.C.;  
 RT "Complete nucleotide sequence and genome organization of bovine  
 RT parvovirus."  
 RL J. Virol. 60:1085-1097(1986).  
 CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M14363; AAB59848.1; -;  
 DR EMBL; M14363; AAB59849.1; -;  
 DR PIR; A26104; VCPVB5.  
 DR HSP; P30129; 4DPV.  
 DR InterPro; IPR001403; Parvo coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 DR Coat protein; Glycoprotein.  
 KW CHAIN 1 673 COAT PROTEIN VP2.  
 FT CHAIN 138 673 COAT PROTEIN VP3.  
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 446 446 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 639 639 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT DOMAIN 163 187 GLY-RICH.  
 SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

## Query Match

Best Local Similarity 17.8%; Score 517; DB 1; Length 673;  
 Matches 143; Conservative 85; Mismatches 208; Indels 134; Gaps 17;

QY 5 GGAPMADNNEGADGVNAGSNHCDSTWLGDRVITSTRTWALPTVNNHLYKQISSASTG 64  
 Db 163 GSGSVGGGGRGSGVSTGGTGTGFSENIVTNTQTFICDKNGHLYKS-EVLNTG 221  
 QY 65 ASNDNHFGYSTPWGDFRNFHCHSPDRQRLNNWNGFPRKLNFKLFIQKVEVTT 124  
 Db 222 DTAHQY-AITTPWSYFNQYSSHPSPNDQHLVNDYERFRPKAMIVRVYNLQIKQIMT 280  
 QY 125 NDGVTTIANN-LTSTVQVSDSEYQLPYVLGSAHQCLPFPFADVPMIPQYGLYL- 179

Db 281 DGAMGTYNNDLTAGMHI FCDGDRHYEYQHPWDDQCMPELNSIWELPQYAYIPAPISV 340  
 QY 180 --NNGSOAVGRS-----SFVLEYFPQMLRTGNNTFSVTFEEVPHSSVAHSQSLDRL 232  
 Db 341 VDNNTTIVTEHLLKGVPLYLENSDHEVLNG-----RI 375  
 QY 233 MNPLIDQYLVLNRTQN-----QSSAQNKKDLFSRGRP-----AGMSVQPKWL 277  
 Db 376 YRIYIQLWRLMDRKQHHI QHASDDVQSTGQKNLLIQRTPKQNKORFQNAALRTSNM 435  
 QY 278 PGCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKODEDEKFFPMG 337  
 Db 436 SGP-----GIARGTHNATLQTSAGALVTWVTNGAD-----VSG 469  
 QY 338 VM-----IFGKE-----SAGASNTALDNVMTDEEBIKATNPVAT 372  
 Db 470 VRAVRGVSTDPYGGQOPESDILLRLRYASASAEQQNPILN-----AAR 515  
 QY 373 ERFGTVAVNFQSSSTDPAQGVHMGALPGVWQDRDVLVQGPWAKI PHTDGHFHPSP 432  
 Db 516 HTFTREARTKLITGNGADGDYKEWMLPNQMMDSAPISRYNPVWVPRVNRKTLTDQ 575  
 QY 433 MGGFGLKNPPQIILK--NTPVPANPPAFAFSATKFAFITQYSTGYVSEIWELEOKENS 490  
 Db 576 DGSIPMSHPPTGTFIKLARIPVPGND-----SFLNIYVTGVSCVVEVERKGT 626  
 QY 491 KRNPEVQYTSNYAKSANVD--FTVDNNGLY 519  
 Db 627 KWRPEYMHs-----ATNMSVDAYTINNAGVY 653

## RESULT 3

COAT\_PAVHB STANDARD; PRT; 781 AA.  
 AC P07299;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 01-APR-1990 (Rel. 14, Last annotation update)  
 DE Probable coat protein VP1.  
 OS Human parvovirus B19.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.  
 OX NCBI\_TaxID=10798;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate AU.  
 RX MEDLINE=86200451; PubMed=3701931;  
 RA Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;  
 RT "Nucleotide sequence and genome organization of human parvovirus B19  
 RT isolated from the serum of a child during aplastic crisis."  
 RL J. Virol. 58:921-936(1986).  
 CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M13178; AAB6867.1; -;  
 DR PIR; A24299; VCPV19.  
 DR InterPro; IPR001403; Parvo coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 781 AA; 86015 MW; 8C6254DBD0576B07 CRC64;  
 Query Match 16.8%; Score 487.5; DB 1; Length 781;



```

Best Local Similarity 27.4%; Pred. No. 1.8e-27;
Matches 143; Conservative 84; Mismatches 239; Indels 55; Gaps 16;

QY 10 ADNNEGADVGNAS--GNWCHDSWLGDRVITSTRTWALPYNNHLYKQISSASTGASN 67
Db 234 AEASTGAGGGGNSVKSWMSEGAIFANSVCTTFSRQFLIPYDPEHHYKVFSPAASCHN 293

QY 68 D-----NHYFCYSTPWGYFDRNPHCHFSRDRWQRLNNNNWGPFRKLNFLNI 117
Db 294 ASGKBAKCTISPIMGYSTFWRYLDNALNLPSPLEFQHLIENYGSIAPDALTWTISEI 353

QY 118 QKVEVT--TNDGVTTIANNLTSTVQVSDSEVQLPVLGSAHQGLCPPPPADVFIMPQKG 175
Db 354 AVKDVTDKGGGV-QVTDSTTGRCLMLVDHEKYPPYVLGGQDITLAPELPIWVYPPQYA 412

QY 176 YLTLAN-GSQAVG-----RSSPYCLEYPPSQMLRTGNFTFSYTFEEVPEHSSYAH 225
Db 413 YLTGVQVNTQIGSGSKLASBASAYVLEHSHFSQLLGTGGTASMSYKFPVPPENLEG 472

QY 226 SOSDLRMLNPLIDQYLYLNRTQNSGSAQNKLDFSRGSPAGMSVQPKNWLPGPCYRQ 285
Db 473 SQHFYEMYNPL---YGRLGVPDTLGGDPKFRSL-----THEDHAIQPNFPGPLVNSV 524

QY 286 RVSKYKTDNNNSFTWAGSKYNLNGRESIINPG-TAMASHKDDDEKDFPMSGVMIFGKE 344
Db 525 STKEGSSNTGAKALITGLSTGTSQNTIRISLRPGVPVQPYHHWDTKYVYTGAINAISHG 584

QY 345 SAGASNTALDNV-----WIDEERIKATNPVATERFGVAVNFSSSTDPTATGDVHAM 397
Db 585 TVG---NAEDKEYQVQGVFPNEKEKQLOGLNMHTY-----FPNKQTOYTOIE-R 634

QY 398 GALPGMVQDRDYLQGPVIAKIPHTDGHFHS-PLMGFGGLKNPPQILIKNTVPVNPAN 456
Db 635 PLMGVSVNRRALHYESQLWSKIPNLDDSKTKQFAALGGWGLHQPPQIFLK--ILPQSG 692

QY 457 PAFSFKATKASFTQYSTQGVSEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEI 496
Db 693 PIGGIKSMGITLVQVAVGIMTVTMTFLGPRKATGRWNPQ 733

RESULT 4
COAL_PAVP9 STANDARD; PRT; 729 AA.
AC P33484;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Porcine parvovirus (strain 90HS) (PPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=33725;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89319168; PubMed=2750278;
RA Sakurai M., Nishimori T., Ushimi C., Nakajima H.;
RT "Nucleotide sequence of the capsid protein gene of porcine parvovirus.";
RL Virus Res. 13:79-86(1989).
CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -2- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -3- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
DR HSP; P30129; 4DPV.
DR InterPro: IPR001403; Parvo coat.
DR Pfam: PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 729 COAT PROTEIN VP1.
FT CHAIN 151 729 COAT PROTEIN VP2.
FT CARBOHYD 172 729 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 729 AA; 80938 MW; 92538BFF9A0C78E6 CRC64;

Query Match 8.0%; Score 232.5; DB 1; Length 729;
Best Local Similarity 23.0%; Pred. No. 3.7e-09;
Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;

QY 3 SGGGAPMADNNEGADVGNASGNWH--CDSTWLGDP---RVITSTRTWALPYNNHLYKQ 57
Db 174 SGGGGG-GGGGRGAGGVGVSTGSFNNQTEFQVGLGELVRITAHASRLIHANPEHETVKR 232

QY 58 IS--SASTGAS-----NDNHFCYSTPWGYFDRNPHCHFSRDRWQRLNNNNWGPFRKLN 111
Db 233 IHLVLSSEGVAGVQVDDAHTMTFWSLIDANAGVMFNPADWQLIISNNTEINLVSFE 292

QY 112 FKLFIQVKEVT---TNDGVTTIANNLTSTVQVPSDSBYQLPVLGSAHQGLCPPPPADV 168
Db 293 QEIFNVVLKTTITESATSPTKIYNDLTASLMVALDTNNTLPTTAAAPRSETLGPYPLP 352

QY 169 FMIPOGYL-----TLNNGSQAVGRS-----SPCYCLE-YFPQMLRTGNPN 208
Db 353 TKPTQRYVYLSCTRLNPNPTVTGSSQITDSITQGLHSDINFTYIENAVPIHLRTGDEF 412

QY 209 TFSYTFEEVPHSSYAHQSILDR-----LNNPLI--DOYLYLNRTQNSGSAQN 257
Db 413 STGIYHPTKPL--KLTHSQWQNRSLGPPKLLTEPTTEGDOHPTLPAANTKRGYHTM 470

QY 258 DLLFSGSPAGMSVOPKNLPGPCVQRQVRKTKTDNNNSNFTMTGASKYNLNGRESIN 317
Db 471 NNSYTEAT---AIRP-----AQGVNTYPMNFEYSNGGPF-----LT 504

QY 318 P--GTAMASHKDDDEKDFPMSGVMIFGKESAGSANTALDNVMTDEEIKATNP----- 369
Db 505 PIVPTADTYNDDE---PNGAIRFTMGYQHGLTTS-----SOELERYTFNPQSKCR 554

QY 370 VATERFGTVA-VNFSSS-----TDPATG--DVHAMGALP-----GMWQDRDYLQGP 415
Db 555 APKQFNOQAPLNLNTNNGILLPSDPIGGKPNHFMFTLNTYGLPTALNTAPFPNGQ 614

QY 416 IWAKIPHTD--GHFHPSPLMGGFGLK-NPPPQILIKNTFPANPPAFBSA-TKPASFITQ 471
Db 615 IWDKELDTDLKPLH---VTAPFVCKNPPGOLFVKIAP---NLTDFFNADSPQPRIT 668

QY 472 YSTQGVSEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEIWEI 526
Db 669 YSNFWMKGTLTFTAKRSMNNWNPICQHTT-----TAENIGNYI-PTNIG 712

RESULT 5
COAL_PAVPN STANDARD; PRT; 729 AA.
AC P18546; P22964; Q89816;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Porcine parvovirus (strain NADL-2) (PPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10797;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90085785; PubMed=2596019;
RA Vasudevacharya J., Basak S., Srinivas R.V., Compans R.W.;
RT "Nucleotide sequence analysis of the capsid genes and the right-hand
RT terminal palindromic of porcine parvovirus, strain NADL-2.";
RL Virology 173:368-377(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021005; PubMed=2219713;

```

RA Vasudevacharya J., Basak S., Srinivas R.V., Compans R.W.;  
 RT "The complete nucleotide sequence of an infectious clone of porcine  
 RL parvovirus, strain NADL-2.";  
 RN Virology 178:611-616(1990).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90010964; PubMed=2794971;  
 RA Ranz A.I., Mancius J.J., Diaz-Aroca E., Casal J.I.;  
 RT "Porcine parvovirus: DNA sequence and genome organization.";  
 RL J. Gen. Virol. 70:2541-2553(1989).  
 CC - SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
 CC VPI, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.  
 CC - MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.  
 CC - SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; M32787; AAA46917.1; -  
 DR EMBL; M32787; AAA46918.1; -  
 DR EMBL; M38367; AAA46919.1; -  
 DR EMBL; M38367; AAA46921.1; -  
 DR EMBL; D00623; BAA00502.1; -  
 DR PIR; B33302; VCPVPP.  
 DR PIR; B33743; VCPVNA.  
 DR HSP; P30129; 4DPV.  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CHAIN 1 729  
 FT COAT PROTEIN VPI.  
 FT CHAIN 151 729  
 FT COAT PROTEIN VP2.  
 FT CARBOHYD 172 172  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 198 198  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 282 282  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 330 330  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 433 433  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 471 471  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 573 573  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 604 604  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 651 651  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT DOMAIN 175 190  
 FT GLY-RICH.  
 FT CONFLICT 10 10  
 FT G -> GKGSFKGVAVIQLIIFYITG (IN REF. 3).  
 FT CONFLICT 16 18  
 FT YKY -> TI (IN REF. 3).  
 FT CONFLICT 56 56  
 FT Y -> T (IN REF. 3).  
 FT CONFLICT 164 164  
 FT G -> A (IN REF. 3).  
 FT CONFLICT 195 195  
 FT T -> S (IN REF. 3).  
 FT CONFLICT 242 242  
 FT V -> S (IN REF. 3).  
 FT CONFLICT 294 294  
 FT E -> A (IN REF. 3).  
 FT CONFLICT 705 705  
 FT N -> K (IN REF. 3).  
 FT CONFLICT 715 715  
 FT R -> K (IN REF. 3).  
 SQ SEQUENCE 729 AA; 80946 MW; EF816E246C80DB42 CRC64;  
 Query Match 8.0%; Score 232.5; DB 1; Length 729;  
 Best Local Similarity 23.0%; Pred. No. 3.7e-09;  
 Matches 138; Conservative 85; Mismatches 240; Indels 137; Gaps 32;  
 QY 3 SGGGAPADNNEGADGVGNASG--NWHCDSTWLGD---RVITSTRTWALPTVNNHLYKQ 57  
 DB 174 SGGGGG--GGGGRGAGGVSTGFNNQTEFYGLGELVRIATAHASLIHLNMPHEHYTKR 232  
 QY 58 IS--SASTGAS----NDNHVFGYSTPWGYDFNFRHCFHSPROWQRLINNMWFRPKRLN 111  
 DB 233 IHLVNESGVAGQVQDDAHTQMTWPSLLIDANAGWGFNPADWQLISNNMTINLVSEE 292  
 QY 112 FKLFNIVQKVT---TNDGVTTIANNLTSTVQVFSSEVQLPYVLSAHQGCCLPPFPADV 168  
 DB 293 QEIFNVVLKTTITESATSPPTKIYNNDLTASLWALDNTNLTPTPAAPRSETLGFYFWLP 552

QY 169 FMIPQYGYL-----TLNNGSQAVGRS-----SFYCLR-YFSPQMLRTGNP 208  
 DB 353 TKPQYRYLYSCIRNLNPPTVTGQQOITDSIQTGLHSDIMFYTIENAVPIHLRLTGDEF 412  
 QY 209 -TFSYTEEVPPHSSYAHSSQSLDR-----LNPPLI--DOYLYLNRNTQNSGSAQNK 257  
 DB 413 STGIYHFDTKPL--KLTHSWQTNRSLSGLPKLLTEPTTEGDOHGTLPAAINTRKGYHQTI 470  
 QY 258 DLLFSRSGPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIIN 317  
 DB 471 NNSYTEAT---AIRP-----AQGVNTPYMNFEYSNGGPF-----LT 504  
 QY 318 P--GTAMASHKDDKDFPMGSMWIFGKESAGASNTALD---NMVITDEEIK--ATNP- 369  
 DB 505 PIVPTADTYNDDEPN-----GAIRFTMDYQHGHLTSSQSLERYTFNPQ 549  
 QY 370 ----VATERGTVA--VNFOSSS-----TDPATG--DVHANGALP-----GMVQQRDV 410  
 DB 550 SKGRAPKQQNQQAPLNLENTNGTLLPSDPICGKSNHFMWTLNTYGLTALNNTAPV 609  
 QY 411 YLGQPIWAKIPHTD--GHFHPSPLMGGGLK--NPPQIILIKNTVPVNPANPPAPSA--TKFA 466  
 DB 610 FPNQGIWDKELDTLKPRLH---VTAPFVCKNPNPGQLFVKIAP---NLTFDFNADSPQ 663  
 QY 467 SFITQYSTGQSVSEIENWELCKENSKRNPEVQYTSNYAKSANVDFTVDDNGLYTEPRPIG 526  
 DB 664 PRIITYSNFWKGTLTFTAKRSMNMNMQOQHTT-----TAENIGNYI--PTNIG 712  
 RESULT 6  
 COAL\_PAVPK STANDARD; PRT; 729 AA.  
 ID COAL\_PAVPK AC P52501;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 OS Porcine protein VPI [Contains: Coat protein VP2].  
 OS Porcine parvovirus (strain Kresse) (PPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=73487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96183900; PubMed=8642680;  
 RA Bergeron J., Hebert B., Tijssen B.;  
 RT "Genome organization of the Kresse strain of porcine parvovirus:  
 RT identification of the allotropic determinant and comparison with  
 RT those of NADL-2 and field isolates.";  
 RL J. Virol. 70:2508-2515(1996).  
 CC - SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
 CC VPI, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.  
 CC - MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.  
 CC - SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; U44978; AAC40230.1; -  
 DR EMBL; U44978; AAC40231.1; -  
 DR HSP; P30129; 4DPV.  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CHAIN 1 729  
 FT COAT PROTEIN VPI.  
 FT CHAIN 151 729  
 FT COAT PROTEIN VP2.  
 FT CARBOHYD 172 172  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 198 198  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 729 AA; 80835 MW; B6345BFA0568A1F6 CRC64;  
  
Query Match 7.9%; Score 230.5; DB 1; Length 729;  
Best Local Similarity 23.0%; Pred. No. 5.2e-09;  
Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;  
  
QY 3 SGGGAPMADNEGADGVGNAGNWH--CDSTWLGCD---RVITSTRTWALPTYNHLYKQ 57  
DB 174 SGGGGG-GGGGRGAGGVSTGSGFNQTEPQYLGELVRAITAHASRLIHLNPEHEITYKR 232  
  
QY 58 IS--SASTGAS-----NDNHVFGYSTPWGDFNRFCHFSRPMQBLINNNWGFPRKRLN 111  
DB 233 IHLNSESAGVAGVQVDDAHTQMTWPSLIDANAWGFNPADWQLISNNMTINLVSP 292  
  
QY 112 FKLPIQVKEVT---TNDGVTTIANLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADV 168  
DB 293 QELPNVVLKTTESATSPPTKIYNNDLTASLWALDNTNLTPTPAAPRSETLGFYFWLP 352  
  
QY 169 FMIPQYCYL-----TLNNGSOAVGRS-----SFYCLE-YPPSOMLRGNNP 208  
DB 353 TKPTQRYVYLSCTENLNPPTTQSQQITDSIQTGLHSDIMFTYENAVPIHLRLTGDEF 412  
  
QY 209 -TPSYTFEEVPHSSVAHQSGLR-----LMNPLI--DOYLYLNTQNGSQAQNK 257  
DB 413 STGIYHFDTKPL--KLTHSQTNRSGLPLPKLLTEPTTEGQDHPGTLPAANTRKGYHTI 470  
  
QY 258 DLLPSRGSPAGMSVQPKWMLPGCYQQRVSKTKTDNNNSNFTWTSKYNLNGRESIIN 317  
DB 471 NNSYTRAT-----AIRP-----AQGYNTPYNFEYSNGGPP-----LT 504  
  
QY 318 P--GTAMASHKDEKFFPMGSGVMI FQKESAGASNTALDNVIMITDEEIKATNP----- 369  
DB 505 PIVPTADTQVNDDE---PNGAIRFTMGYQHGLTTS-----SQELRYTFNPSQKCR 554  
  
QY 370 VATERFGTVA-VNFQSS-----TPATG--DVHANGALP-----GMWQDRDVLQGP 415  
DB 555 APRQFNOQAPLNTNTGTLPSDIPGCKPMHFMNTLNTYGLTALNTAPVPPNQ 614  
  
QY 416 IWAKIPTHD--GHFHSPLMGFGGLK-NPPQILLIKNTVPANPPAERSA-TKFASFIQ 471  
DB 615 IWDKELDTLKLPLH---VTAPVCNKNPPGQVFKIAP---NLTDGFNADSPQQRPIIT 668  
  
QY 472 YSTGVQSVRIEWELOKENSKRMPVEQYTSNYSKASANVDFTVDNNGLYTEPRPIG 526  
DB 669 YSNFNNWKTTLTFAKWRSSNMNPIQGHIT-----TAENIGNYI-PTNIG 712

RESULT 7  
COAT\_PAVL3  
ID COAT\_PAVL3 STANDARD; PRT; 587 AA.  
AC P36310;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coated protein VPI [Contains: Coated protein VP2].  
OS Parvovirus LuIII.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=35339;  
RN [1]  
RX MEDLINE=93297126; PubMed=8517025;  
RA Diffeot N., Chen K.C., Bates R.C., Lederma M.;  
RT "The complete nucleotide sequence of parvovirus LuIII and  
RT localization of a unique sequence possibly responsible for its  
RT encapsidation pattern."  
RL Virology 192:339-345(1993).

CC CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC CC or send an email to license@isb-sib.ch).  
CC CC -----  
CC CC EMBL; M81888; -; NOT ANNOTATED\_CDS.  
CC CC PIR; B44276; B44276.  
CC CC HSSP; P07302; LMW.  
CC CC InterPro: IPR001403; Parvo coat.  
CC CC Pfam: PF00740; Parvo\_coat; 1.  
CC CC Coated protein; Glycoprotein.  
CC CC CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CC SQ SEQUENCE 587 AA; 65429 MW; 523E6B9CB2EBB74 CRC64;  
  
Query Match 7.9%; Score 229; DB 1; Length 587;  
Best Local Similarity 22.7%; Pred. No. 4.9e-09;  
Matches 133; Conservative 74; Mismatches 228; Indels 150; Gaps 28;  
  
QY 3 SGGGAPMADNEGADGVGNAGNWHCDSD---TWLGD---RVITSTRTWAL---PTYNN--- 52  
DB 30 SGGG-----GSGGGGVSTGSDYDQTHYKPLGSGWVEITAYSTRMVLNMPKSENYCR 83  
  
QY 53 ---HLYKQISASTGASNDNHVFGYSTPWGDFNRFCHFSRPMQBLINNNWGFPRK 109  
DB 84 VRVHTNDTGTASHMADDAHEQIW-TWLSLVANAWGVFQPSDWQVYSNNMIHLNLS 142  
  
QY 110 LNFKLPIQVKEVT-TNDGVTTIA---NNLSTVQVFSDEYQLPYVLGSAHQGCL----- 161  
DB 143 LDQELFNVIKTVTQNTGAEAIKVVYNNDLTAAMVVALDSNNILPYTPALDNDQETLGFYP 202  
  
QY 162 --PPFPA-----DVFMIPOQYGLTVLNGSQAVGRSFFCYCLEVFP--QMLRTGN 206  
DB 203 WKPTIPSPYRYFSCDRNLSTVYKDEAGTITDTMGLASGLNSQFFTIENQRLNLTG 262  
  
QY 207 NF-TPSYTFEEVPHSSVAHQSGLRNLNPLIDQ-----LYLNLRTQN- 249  
DB 263 EVATGYTFDTEPIRLTHWTQNRHLGPPQITELPSSDTANATLTARGYRSLTQIQGR 322  
  
QY 250 -----QSGSAQNKD--LFSRGSAGMSVQPKWLPD-----PCYQQR 286  
DB 323 NDVTEATRVPAQVGFQCFQPHDNFETSRAGPFKVPVVPADITQGLDHDANGSLRYTYDKH 382  
  
QY 287 VSKTKTDNNNSNFTWTSKYNLNGRESIINPGTAMASHKDEKFFPMGSGVMI FQKESA 346  
DB 383 QGSWASQNNKORYTW-DAVNYD-SGR----- 406  
  
QY 347 GASNTALDNVIMITDEEIKATNPVATER---FGTVAVNFQS---SSTDPATGVDHANGALP 401  
DB 407 WTNNCFIQSVPTSEPN---ANQILTNRLNLAGTKTDIHTNAPNSVGPLTAPPH----- 457  
  
QY 402 GMWQDRDVLQGPPIWAK---IPHTDGHFHSPLMGFGGLKKNPPQILLIKNTP--VPANP 456  
DB 458 -----PAPIYPGQGIWDKELDLKPLHRTQAPFY---CKNAPGQLLVRLAPNLTDQYD 509  
  
QY 457 PAERSATKFPASFITQYTSNGVSVETWELELOKENSKRMPVEQYTS 501  
DB 510 PNSSNLRIYVTGTFWKGKUTLKAK---MRPNA-TWNPVPOISA 550

```

RESULT 8
COAT_PAVHH STANDARD; PRT; 722 AA.
ID COAT_PAVHH
AC P03136;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Hamster parvovirus H1.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10799;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83112183; PubMed=6823009;
RA Rhode S.L. III, Paradiso P.R.;
RT "Parvovirus genome: nucleotide sequence of H-1 and mapping of its
   genes by hybrid-arrested translation.";
RL J. Virol. 45:173-184(1983).
CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC
DR EMBL; X01457; CAB57285.1; ALT_SEQ.
DR PIR; A03699; VCPV2.
DR HSSP; P07302; IMVM.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 722 COAT PROTEIN VP1.
FT CHAIN 131 722 COAT PROTEIN VP2.
FT CARBOHYD 178 178 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 647 647 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DOMAIN 155 170 GLY-RICH.
FT SEQUENCE 722 AA; 79737 MW; 68678391AA5DC31 CRC64;

Query Match 7.8%; Score 226; DB 1; Length 722;
Best Local Similarity 21.7%; Pred. No. 1.1e-08;
Matches 128; Conservative 83; Mismatches 234; Indels 146; Gaps 26;

QY 2 ASGGAPMADNNEGADGVGNAGNWHCDSTW--LGDRVITSTRTWAL----- 47
DB 153 ADGGGS--SGGGGGGGGIGVSTGYDQTYKFLGDGWEITAHAGRLHLGMPSPSENYC 211
QY 48 --PTYNHLYKQISSASTGASNDNHFGYGTPTWGYFDNRFHCFSPRDMQRLNNMGF 105
DB 212 RVTVHNNQTTGHTKVGKNWAYDTHQIWI--TPSLVDANAGWVGFQPSDQPTQNSMESL 270
QY 106 RPKEINPKLNIQVKEVTT-----NDGVTTIANNLSTVQVPSDSEYQLPYVLGSAHQGC 160
DB 271 NLDSLSQELFNWVVKVTVEQQAGQDAIKVYNNDLTACMVMALDSNNILPYTPAAQTS 330
QY 161 L-----PPPAD---VPMTPGYLYTLNNGSQ-----AVG-----RSSFCYCLE-YFPS 199
DB 331 LGFYFWKPTAPARYRYFFMPRQLSVTSSNAEGTQITDTIGBPAQLNSQFFTIENLFI 390
QY 200 QMLRTGNFTF-SYTFEEVPPHSSYA-----HSQSLRLMN-PLIDQYLYLNARTONQSG 252
DB 391 TLLRTGDEFTTGTIFNTPDKLTHWTQWHLACLQIGITDLFSDTATASLTANGDRFG 450
QY 253 SAQNKDLFF-----SRGSPAGMSVQPKWLPGPCYRQORVSK 289
DB 451 STQTONVYVTEALRTPAQIGFMQPHDNFEANRGGFKVPVVP-----LDI 497
QY 290 TKTDNNNSNFTWTGASKYNLNGR--ESIINPGTA-----MASHKDEDEKFFPMMSG 337

RESULT 9
COAT_MUMIV STANDARD; PRT; 716 AA.
ID COAT_MUMIV
AC P03137;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Murine minute virus (Murine parvovirus).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83143341; PubMed=6298737;
RA Astell C.R., Thomson M., Merchinsky M., Ward D.C.;
RT "The complete DNA sequence of minute virus of mice, an autonomous
   parvovirus.";
RL Nucleic Acids Res. 11:999-1018(1983).
CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC
DR EMBL; V01115; CAA24310.1; ALT_SEQ.
DR PIR; A03700; VCPV2M.
DR HSSP; P07302; IMVM.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 716 COAT PROTEIN VP1.
FT CHAIN 131 716 COAT PROTEIN VP2.
FT CARBOHYD 179 179 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DOMAIN 156 171 GLY-RICH.
FT SEQUENCE 716 AA; 78707 MW; 6A7229A91161F4C6 CRC64;

Query Match 6.8%; Score 197; DB 1; Length 716;
Best Local Similarity 20.0%; Pred. No. 1.3e-06;
Matches 126; Conservative 88; Mismatches 241; Indels 174; Gaps 28;

QY 3 SGGGAPMADNNEGADGVGNAGNWHCDSS--TWLGD---RVITSTRTWALPTNNHLYKQ 57
DB 160 SGGG-----GSGGGGVGVSTGSDYNQTHYRFLGDGWEITATLRLVHLNMPKSENYCR 213
QY 58 I-----SSASTGASNDNHFGYGTPTWGYFDNRFHCFSPRDMQRLNNMGFPKRL 110
DB 214 IRVHNTDTSVKGNMAKODAEHQIWTWPSLVANAGWVWLQPSDMQYICNTMSQLMLVSL 273
QY 111 NPKLFNIQVKEVTTND-----GVTTIANNLSTVQVPSDSEYQLPYVLGSAHQGCCLPPPPA 166

```

```
Db 274 DOEIFNVLTQVTEQDLGGQAIKIYNNDLTACMVAVDSDNNILFYTPAANSMETLGFYEW 333
Qy 167 DVFMIPQYGV-----LTLNNGQAV-----GRSSPYCLEYEPS-QMLRTGNN 207
Db 334 KPTIASPYRYFYCVDRDLSTVYENQSGTVEHNVNMGPKGIPQFTTIENTQQITLLRTGDE 393
Qy 208 F-TFSYTFEEVPPHSSYAHSSQSLDRLMN--PLIDQYLYLNRNTQNSGSAQNKKDLIFSRS 264
Db 394 FATGTYYFDNSV--KLTHWTQNRQLGQPLLSTF-----PEADTDAGT-----LTAQG 441
Qy 265 SPAGMSVQPKNWL-----PQFCYRQQRVSKTKTDNNNSNFT 300
Db 442 SRHGTTQMGVNVYSEAIRTPAQVGFQCPHNDPEASRAGP-FAAPKVPADITQGVDKKAN 500
Qy 301 WTGASKYNLNGRESIINPG-----TAMASHKDDDEKFP-----PMSGVMILF 341
Db 501 GSVRYSGYQKHGHNASHGAPRYTWDSTFSFGSDRTKDGFIQASPLVVPPLNGI----- 557
Qy 342 GKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFOS--SSTDPAATGDVHAMGA 399
Db 558 -----LTNANPIGTKN-----DIHFSNVFNSYGLTAFSH----- 587
Qy 400 LPGMWQDRDVLQGIWAK---IPHTDGHFHPSPLMGGFGLKNPPPIILIKNTP--VPA 454
Db 588 -----PSPVYPQGIWDEKLEHKLPHLITAPFV---CKNNAFGQMLVRLGPNLTDQ 637
Qy 455 NPPAEPFSAFKAFITQYSTGVSGVSEIWELOKNSKRWNPVEQVTSYNAKSANVDFTVD 514
Db 638 YDENGATLSRIYTYGTFYFKGKLTWRAKURA-----NTWNVPVYQVSAE-----D 682
Qy 515 NNGLY---TEPRPIGT-----RYLTRPL 534
Db 683 NGNSYMSVTKWLPTATGNMQSVPLITRPV 711

RESULT 10
COAT_MUMIM STANDARD; PRT; 718 AA.
AC P07302; Q9WMH2; Q9WMH3;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Coat protein VP1 (Contains: Coat protein VP2).
OS Murine minute virus (strain MVM1) (Murine parvovirus).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10795;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=86115415; PubMed=3502703;
RA Astell C.R., Gardiner E.M., Tattersall P.;
RT "DNA sequence of the lymphotropic variant of minute virus of mice,
RT MVM(1), and comparison with the DNA sequence of the fibrotropic
RT prototype strain.";
RL J. Virol. 57:656-669(1986).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=85242059; PubMed=3855242;
RA Sahli R., McMaster G.K., Hitt B.;
RT "DNA sequence comparison between two tissue-specific variants of the
RT autonomous parvovirus, minute virus of mice.";
RL Nucleic Acids Res. 13:3617-3633(1985).
RN [3]
X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718.
RA Llamas-Saiz A.L., Agbandje-Mckenna M., Wikoff W.R., Bratton J.,
RA Tattersall P., Rosemann M.G.;
RT "Structure determination of Minute Virus of mice.";
RL Acta Crystallogr. D 53:93-100(1997).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02481; CAB46507.1; -.
DR EMBL; X02481; CAB46508.1; -.
DR EMBL; M12032; AAA69569.1; ALT_INIT.
DR PIR; B23008; VCPVIM
DR PDB; 1MVM; 25-FEB-98.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein; 3D-structure.
FT CHAIN 1 718
FT COAT PROTEIN VP1.
FT CHAIN 132 718
FT CARBOHYD 180 180
FT CARBOHYD 180 180
FT CARBOHYD 219 219
FT CARBOHYD 502 502
FT CARBOHYD 635 635
FT CARBOHYD 671 671
FT DOMAIN 157 172
FT CONFLICT 144 144
FT STRAND 182 186
FT STRAND 191 204
FT STRAND 213 217
FT TURN 220 221
FT TURN 228 231
FT STRAND 235 245
FT TURN 251 253
FT HELIX 256 265
FT STRAND 266 266
FT STRAND 268 269
FT STRAND 274 289
FT STRAND 296 301
FT TURN 303 304
FT STRAND 307 307
FT STRAND 309 312
FT HELIX 322 325
FT TURN 326 326
FT TURN 333 334
FT STRAND 337 338
FT STRAND 340 345
FT STRAND 351 351
FT TURN 357 358
FT STRAND 367 367
FT TURN 373 375
FT HELIX 381 384
FT STRAND 390 390
FT STRAND 396 397
FT STRAND 408 409
FT STRAND 428 428
FT STRAND 438 438
FT HELIX 442 444
FT TURN 458 460
FT STRAND 466 467
FT TURN 473 474
FT STRAND 477 477
FT STRAND 484 484
FT TURN 494 496
FT TURN 499 503
FT STRAND 509 509
FT TURN 512 513
FT STRAND 525 525
FT HELIX 534 536
FT STRAND 546 546
FT TURN 557 558
FT STRAND 559 559
FT TURN 575 577
FT STRAND 583 584
FT STRAND 585 586
FT STRAND 600 600
FT STRAND 618 618
FT STRAND 627 630
FT A -> G (IN REP. 2).
```



661 TYSDFWFKGLVFKAKLRASHWTNPIQMSIN-----VDNQFNYL-PNNIG 705

RESULT 12

COAT\_PPV

ID COAT\_PPV STANDARD; PRT; 727 AA.

AC P04864; Q65112;

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Coat protein VP1 [Contains: Coat protein VP2].

OS Feline panleukopenia virus (FPV).

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.

OX NCBI\_TaxID=10786;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95265017; PubMed=2991581;

RA Carlson J., Kushnir K., Maxwell I., Maxwell S., Hahn W.;

RT "Cloning and sequence of DNA encoding structural proteins of the

RT autonomous parvovirus feline panleukopenia virus.";

RL J. Virol. 55:574-587(1985).

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M10824; AAA47161.1; -

DR EMBL; M10824; AAA47162.1; -

DR PIR; A03701; VCPV1F.

DR HSP; P30129; 4Dpv

DR InterPro; IPR001403; Parvo coat.

DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein.

FT CHAIN 1 727

FT CHAIN 144 727

FT CARBOHYD 168 168

FT CARBOHYD 190 190

FT CARBOHYD 207 207

FT CARBOHYD 323 323

FT CARBOHYD 586 586

FT CARBOHYD 648 648

FT CARBOHYD 660 660

FT DOMAIN 165 180

FT GLY-RICH.

FT SEQUENCE 727 AA; 80344 MW; 770E0F6B554C0C6E CRC64;

Query Match 6.7%; Score 194; DB 1; Length 727;

Best Local Similarity 20.7%; Pred. No. 2.2e-06;

Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;

QY 5 GGAPMADNNEGAGVGNGAGNHCDSWGLGDRVITTS-----RTWALPTVNN- 52

DB 153 GGQP-AVRNERATGSGNGSGGGGGG--GGVGISTGTGTFNNQTEFKPLNGWVEITANS 209

QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPWGYPDFNRFCHFSRWD 95

DB 210 RLVLNMPESNTKRVVNNNDKAVKGNMADDTTHVQIVTFWLSVDANAWGVWFPNGDW 269

QY 96 QRLINNMGPRKPLNFKLFNIQVKEVT---TNDGVTTIANLITSTVQVPSDSEYQLPYV 152

DB 270 QLIVNTMSLHLVSFEQIEFNVLKTVSESATPPTKVYNNDLTASLMVALDSNNMTMPT 329

QY 153 LGSAGHOCGL-----PPPPA-----DVFMTPQGYLTNNGSQAVER 189

DB 330 PAAMRSETLGFYPMKPTIPTFWRYFQWDRTLIPSH-----TGTSGTPTNIYHGDPDD 383

QY 190 -SFYCLR-YFPQMLRTGNPF---TF-----SYTPEE-----VPHSSVAHSQSL 229

DB 384 VQYTIENSVPVHLRTGDEFGATGFFFDCKPCLRLTHWTQNRALGLPPLFNSLPQSEGA 443

QY 230 DR-----LMNPLIDOYL--YYLNRTQNS----- 251

DB 444 TNFGDIGVQODKRGVGTQMGNTDYITEATIMRPAEYGVSAPIYSFEASTGPKPIPAAG 503

QY 252 -GSAQNKDLFLSRSGPAGMSVOPKWLPGFCY---RQORVSKTKTNNNSNFNTWTGASKY 307

DB 504 RGGAQTDENQAADGDR-----YAFGRHQGKTKTTTGETPERFTY----- 543

QY 308 NLNGRESINPGTAMASHKDDKFFPMGVMIFGKESAGASNTALD-NVMITDEBEIKA 366

DB 544 -----IAHQDT-----GRYPAGDMTQINFNFLPVTNDNVLLP 575

QY 367 TNVATERGTVAIVNFOSSTDPATGDVHAMGALPGMWQDRDVLVQLGPTWAKIPTHDDGH 426

DB 576 TDPIG---GKTGINY--TNIFTYGLTALNNVP-----PVYPNGQIWDKEFDD-- 620

QY 427 FHPSPMLMGFGFK-----NPPQLIKNTVPAN---PPAFSFAFKFASFIT 470

DB 621 -----LKPRLLHVNAPFVQNNCPGQLFVKVAPNLITNEYDPPDANMSR---IV 665

QY 471 QYSTGQSVSEIWELOKSKENNEVEQYTSYNAKSAVDFVTDNGLTEPRPIG 526

DB 666 TYSDFWFKGLVFKAKLRASHWTNPIQMSIN-----VDNQFNYL-PNNIG 710

RESULT 13

COAT\_PAVC2

ID COAT\_PAVC2 STANDARD; PRT; 584 AA.

AC F30129;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 28-PEB-2003 (Rel. 41, Last annotation update)

DE Coat protein VP2.

OS Canine parvovirus (type 2 / strain A72) (CPV).

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.

OX NCBI\_TaxID=31597;

RN [1]

RP X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).

RX MEDLINE=91173306; PubMed=2006420;

RA Teao J., Chapman M.S., Agbandje M., Keller W., Smith K., Wu H.,

RA Luo M., Smith T.J., Rosemann M.G., Compans R.W., Parish C.R.;

RT "The three-dimensional structure of canine parvovirus and its

RT functional implications.";

RL Science 251:1456-1464(1991).

CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,

CC OF A COMBINATION OF VP2, VP3, AND SOME VP1.

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC PDB; 4DFV; 01-APR-97.

DR PDB; 11JS; 23-DEC-96.

DR InterPro; IPR001403; Parvo coat.

DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein; 3D-structure.

FT CARBOHYD 25 25

FT CARBOHYD 47 47

FT CARBOHYD 64 64

FT CARBOHYD 180 180

FT CARBOHYD 443 443

FT CARBOHYD 505 505

FT CARBOHYD 517 517

FT TURN 38 39

FT STRAND 49 53

FT STRAND 58 64

FT STRAND 67 71

FT STRAND 80 84

FT TURN 87 89

FT TURN 102 112

FT TURN 118 120

FT TURN 123 132

FT STRAND 133 156

FT STRAND 162 167

FT TURN 169 170

FT STRAND 173 177

```

FT TURN 179 180
FT TURN 188 191
FT TURN 199 200
FT STRAND 203 203
FT STRAND 206 211
FT STRAND 215 218
FT TURN 231 235
FT TURN 238 239
FT STRAND 246 249
FT STRAND 261 262
FT STRAND 266 267
FT STRAND 270 270
FT STRAND 273 274
FT TURN 300 301
FT TURN 310 312
FT STRAND 317 317
FT TURN 327 329
FT STRAND 330 330
FT STRAND 335 336
FT STRAND 339 339
FT STRAND 346 347
FT STRAND 352 353
FT TURN 365 366
FT TURN 370 374
FT STRAND 378 380
FT TURN 388 389
FT STRAND 396 398
FT TURN 410 411
FT STRAND 415 416
FT TURN 418 419
FT HELIX 426 428
FT STRAND 429 430
FT TURN 445 447
FT STRAND 450 450
FT TURN 453 454
FT STRAND 455 456
FT STRAND 464 464
FT TURN 465 466
FT STRAND 470 470
FT STRAND 488 488
FT STRAND 490 490
FT STRAND 498 502
FT STRAND 506 506
FT TURN 512 513
FT STRAND 520 520
FT STRAND 523 539
FT TURN 555 557
FT STRAND 558 560
FT HELIX 565 566
FT TURN 574 574
SQ SEQUENCE 584 AA; 64686 MW; 6E4DADA5AEBF9D8C CRC64;

Query Match
Best Local Similarity 6.5%; Score 190; DB 1; Length 584;
Matches 130; Conservative 82; Mismatches 229; Indels 198; Gaps 30;

QY 5 GGAPMADNNEGADGVGNAGNHCDSWLGDRVITST-----RTWALPTYNN- 52
DB 10 GGOP-AVRNERATSGNGSGGGGGG--GGVGISTGTFNNQTEFKFLENGWVYITANS 66
QY 53 ---HL-----YKQI-----SSASTGASNDNHVFGYSTPMGYDFNRFCHFSPRDW 95
DB 67 RLVLHAMPESENRYRVVANNMKTAVNGNMLDDIHAIEIVTWSLVDANAWGWFNPGDW 126
QY 96 ORLNNWGRPRKRLNFKLFNIOVKET---TNDGVTTIANNLTSTVQVFSSEYOLPVV 152
DB 127 QLIVNTWSELHLVSFQEIENNVLVKTVVSATQPTKVYVNDLTASLWALDSNNTMPT 186
QY 153 LGSAGHGCCL-----PPPPA-----DVFMIPQYGYLTLLNNGSQAVGRS----- 189
DB 187 PAAMRSETLGFYWKPTIPTFPRYRYFQWDRTLIPSH-----TGTSGTPTNIYVHGDDPD 240

```

```

QY 190 -SFYCLE-YFPQMLRTGNMF-TFSYTFEVPFPHSSVAHSSQSLDRLMN--PLIDQVLYYL 244
DB 241 VQFYTIENSVPVHLRTGDEFATGTFDFCKP--CRLTHTWQTNRALGLPP-----FL 291
QY 245 NRTQNSGSAQNDLLFSRGSAPAGMSVQPKNWLPGCYRQORVSKTKTDNNNSNF----- 299
DB 292 NSLPQSEGA TNFGDI-----GV-----QODKRGVTQMGNTNYITEAT 329
QY 300 ----TWGASKYNLNGRESIINP-----GTAMASHKDDDEKFFPMGSMVIMFGKESAGASN 350
DB 330 IMRPAEYGVSAFYYSFEASTQGPPTIAAGRGGAQTDENQAADGNPRYAFGRQHKKTT 389
QY 351 TALD-----NMTDEBEIKATNPVATERFGTVANFQ 383
DB 390 TTGETPERFTYIAHQDTGRYPEGDWIQINFNLPVTNDNVLLPTDPIG-----GKTGINY- 444
QY 384 SSSTDPAATGVHAGALPGMVQDRDVLQGPWAKIPHTDGHFHPSPMLGGFGLK----- 439
DB 445 -TNIFTYGLTALNVV-----PVYFNGQIWDKEFTD-----LXPLRH 483
QY 440 -----NPPQILIKNTFPVAN---PPAFSATSATKPFASFTQYSTQGVSVSEIWELOK 487
DB 484 VNAPFVQNNCPGOLFVKVAPNLTNQYDPPDASANMSR-----IVTYSDFWFKGLVFRACL 539
QY 488 ENSKRWNPVOYTSNYAKSANVDFTVDNGLYTEPRPIG 526
DB 540 RASHTWNP IQMSIN-----VDNQFNYY-PSNIG 567

RESULT 14
COAT_FP19
ID COAT_FP19 STANDARD; PRT; 727 AA.
AC P24840;
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Feline panleukopenia virus (strain 193) (FPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=193/70;
RX MEDLINE=91073139; PubMed=2174965;
RA Martyn J.C., Davidson B.E., Studdert M.J.;
RT "Nucleotide sequence of feline panleukopenia virus: comparison with
  canine parvovirus identifies host-specific differences.";
RL J. Gen. Virol. 71:2747-2753(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CU-4;
RX MEDLINE=91272479; PubMed=1647068;
RA Parrish C.R.;
RT "Mapping specific functions in the capsid structure of canine
  parvovirus and feline panleukopenia virus using infectious plasmid
  clones.";
RL Virology 183:195-205(1991).
RC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X55115; CAA38911.1; -.
DR EMBL; M38246; AAC37928.1; -.
DR EMBL; M38246; AAC37929.1; -.
DR PIR; B36608; VCPVFP.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.

```





Db 533 TTGETPERFYIAHODTGRYPEGDWIONINFNLVPTNDONVLLPTDPIG-----GKTGINY- 587  
Qy 384 SSSDTPATGDVHAMGALPGMVWQDRDVLQGPIMAKIPHHTDGHFHPSPLMGGFGLK---- 439  
Db 588 -TNIFNTYGLTALNVP-----PVYPNGQIWDKEFTD-----LKPRLH 626  
Qy 440 -----NPPQPILIKNTVPAN---PPAEFSATKFASFTQYSTQGSVEIEWELQK 487  
Db 627 VNAPFVQNNCPQLFVKVAPNLITNEYDPAASANMSR-----IVTYSDFWWMKGLVFKAKL 682  
Qy 488 ENSKRNPEVQYTSNVAKSANVDFTVDNNGLYTERPIG 526  
Db 683 RASHTWNPQMSIN-----VDNQFNTV-PSNIG 710

Search completed: January 21, 2004, 16:02:14  
Job time : 11.1429 secs

Result No.	Query	Score	Query			ID	Description
			Match	Length	DB		
1	2906	100.0	736	12	Q9WBP8	Q9WBP8 adeno-associ	
2	2884	99.2	736	12	O56137	O56137 adeno-associ	
3	2530	87.1	736	12	O56139	O56139 adeno-associ	
4	2528	87.0	736	12	O65311	O65311 adeno-associ	
5	2481.5	85.4	533	12	O92317	O92317 adeno-associ	
6	2481.5	85.4	598	12	O56653	O56653 adeno-associ	
7	2481.5	85.4	735	12	O56652	O56652 adeno-associ	
8	2430	83.6	737	12	Q8JQO0	Q8JQO0 adeno-associ	
9	2384.5	82.1	738	12	O8JQF8	O8JQF8 adeno-associ	
10	1698	58.4	534	12	O67668	O67668 goose parvo	
11	1698	58.4	587	12	O67667	O67667 goose parvo	
12	1698	58.4	732	12	O67666	O67666 goose parvo	
13	1697	58.4	732	12	O8V395	O8V395 goose parvo	
14	1690.5	58.2	734	12	O41855	O41855 adeno-associ	
15	1690	58.2	732	12	Q83290	Q83290 muscovy duc	
16	1676	57.7	534	12	O65446	O65446 barbarie du	

```
Db 323 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 382
QY 181 NGSQAVGRSSFCYLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 240
Db 383 NGSQAVGRSSFCYLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 442
QY 241 LYYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 300
Db 443 LYYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 502
QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIFGKESAGASNTALDNVMTD 360
Db 443 LYYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 502
QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIFGKESAGASNTALDNVMTD 360
Db 503 WTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIFGKESAGASNTALDNVMTD 562
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 420
Db 563 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 622
QY 421 PHTDGHFHPSPLMGGFGLKPPQIILKNTVPVPAEFSATKFASTFQYSTGQVSVE 480
Db 623 PHTDGHFHPSPLMGGFGLKPPQIILKNTVPVPAEFSATKFASTFQYSTGQVSVE 682
QY 481 IEWELQKENS KRNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534
Db 683 IEWELQKENS KRNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 2
O56137 PRELIMINARY; PRT; 736 AA.
AC O56137
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Capsid protein VP1.
OS Adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
RT (AAV) serotypes other than AAV type 2.";
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028704; AAB95450.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;

Query Match 99.2%; Score 2884; DB 12; Length 736;
Best Local Similarity 99.1%; Pred. No. 3.6e-206;
Matches 529; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASGGAPMADNNEGADGVGNAGNHCDSWLGDRVITSTRTRWALPTYNHLYKQISS 60
Db 203 MASGGAPMADNNEGADGVGNAGNHCDSWLGDRVITSTRTRWALPTYNHLYKQISS 262
QY 61 ASTGASNDNHYFGYSTPWGDFNRPCHFSPRDWORLNNWGRPKRLNFKLNIQVK 120
Db 263 ASTGASNDNHYFGYSTPWGDFNRPCHFSPRDWORLNNWGRPKRLNFKLNIQVK 322
QY 121 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 180
Db 323 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 382
QY 181 NGSQAVGRSSFCYLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 240
```

```
Db 383 NGSQAVGRSSFCYLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 442
QY 241 LYYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 300
Db 443 LYYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 502
QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIFGKESAGASNTALDNVMTD 360
Db 503 WTGASKYNLNGRESIINPGTAMASHKDDKDEKFFPMGSGVMIFGKESAGASNTALDNVMTD 562
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 420
Db 563 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRDVLQGPWAKI 622
QY 421 PHTDGHFHPSPLMGGFGLKPPQIILKNTVPVPAEFSATKFASTFQYSTGQVSVE 480
Db 623 PHTDGHFHPSPLMGGFGLKPPQIILKNTVPVPAEFSATKFASTFQYSTGQVSVE 682
QY 481 IEWELQKENS KRNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534
Db 683 IEWELQKENS KRNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 3
O56139 PRELIMINARY; PRT; 736 AA.
AC O56139
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Capsid protein VP1.
OS Adeno-associated virus 3B.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68742;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
RT (AAV) serotypes other than AAV type 2.";
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028705; AAB95452.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81906 MW; DD52331AD5F0D70F CRC64;

Query Match 87.1%; Score 2530; DB 12; Length 736;
Best Local Similarity 86.0%; Pred. No. 8.2e-180;
Matches 460; Conservative 27; Mismatches 46; Indels 2; Gaps 2;

QY 1 MASGGAPMADNNEGADGVGNAGNHCDSWLGDRVITSTRTRWALPTYNHLYKQISS 60
Db 203 MASGGAPMADNNEGADGVGNAGNHCDSWLGDRVITSTRTRWALPTYNHLYKQISS 262
QY 61 ASTGASNDNHYFGYSTPWGDFNRPCHFSPRDWORLNNWGRPKRLNFKLNIQVK 120
Db 263 QS-GASNDNHYFGYSTPWGDFNRPCHFSPRDWORLNNWGRPKRLNFKLNIQVK 321
QY 121 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 180
Db 322 EVTTNDGVTTIANNLTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 381
QY 181 NGSQAVGRSSFCYLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 240
Db 382 NGSQAVGRSSFCYLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 441
QY 241 LYYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 299
Db 442 LYYLNRTQNSGSAQNKDLLFSRGSAGMSVQPKWLPGPCYRQORVSKTKTDNNNSFT 501
```



```

QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKFFPMGVMIFGKESAGASNTALDNVMTD 360
DB 300 WTGATKYHLNGRDSLVPNPGPAMASHKDDKFFPMGVMIFGKQSEKTNVDIEKVMITD 359
QY 361 EBEIKATNPVATERFGTVAVNFQSSSTDPAIGDVMHMGALPGVMWQDRDVLQGPWAKI 420
DB 360 EBEIRTNVPATEQVGSVSTNLQRNQAAATADVNTQGVLPGMVWQDRDVLQGPWAKI 419
QY 421 PHTDGHFHPSPMLMGFGGLKNPQIILIKNTPVPANPPAPFSAKTSFTQYSTQGVSE 480
DB 420 PHTDGHFHPSPMLMGFGGLKNPQIILIKNTPVPANPPAPFSAKTSFTQYSTQGVSE 479
QY 481 IEWELQKENSKEWNPVEQVTSNYAKSANVDFTVNNGLYTEPRPTGTRYLTRPL 534
DB 480 IEWELQKENSKEWNPVEQVTSNYAKSANVDFTVNNGLYTEPRPTGTRYLTRPL 533

RESULT 6
OS6653 PRELIMINARY; PRT; 598 AA.
AC OS6653;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Major coat protein VP2.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF043303; AAC03780.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein.
FT VARIANT 416 416 D -> N.
FT VARIANT 430 430 T -> N.
FT VARIANT 540 541 QV -> HV.
FT VARIANT 573 573 V -> R.
SQ SEQUENCE 598 AA; 66619 MW; 070811ED9368E934 CRC64;

Query Match 85.4%; Score 2481.5; DB 12; Length 598;
Best Local Similarity 83.3%; Pred. No. 2.5e-176;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGAPMADNNEGADGVGNAGSNWHCDSTWLGDRVITTTSTRTWALPTYNHLYKQISS 60
DB 66 MATGSGAPMADNNEGADGVGNAGSNWHCDSTWLGDRVITTTSTRTWALPTYNHLYKQISS 125
QY 61 ASTGASNDNHFGYSTPMGWYDFNRFCHFSPRDWRQLNNNMGPRPKRLNFKLFIQVK 120
DB 126 QS-GASNDNHFGYSTPMGWYDFNRFCHFSPRDWRQLNNNMGPRPKRLNFKLFIQVK 184
QY 121 EVTTNDGVTTIANNLTSTQVDSSEYQLPYVLGSAHQCLPFPADVFMIPOYGYLTIN 180
DB 185 EVTQNDGVTTIANNLTSTQVDSSEYQLPYVLGSAHQCLPFPADVFMIPOYGYLTIN 244
QY 181 NGSQAVGRSSFCLEYPFSPQMLRTGNFTSYTFEEVPHSSYAHQSQDLRLNPLIDQY 240
DB 245 NGSQAVGRSSFCLEYPFSPQMLRTGNFTSYTFEEVPHSSYAHQSQDLRLNPLIDQY 304

```

```

QY 241 LYLNRNTQNGSSAQNKDLLFSRGSAGMSVQPKQWLPGPYQQRVSKTKTDNNNSFT 300
DB 305 LYLNRNTQNGSSAQNKDLLFSRGSAGMSVQPKQWLPGPYQQRVSKTKTDNNNSFT 364
QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKFFPMGVMIFGKESAGASNTALDNVMTD 360
DB 365 WTGATKYHLNGRDSLVPNPGPAMASHKDDKFFPMGVMIFGKQSEKTNVDIEKVMITD 424
QY 361 EBEIKATNPVATERFGTVAVNFQSSSTDPAIGDVMHMGALPGVMWQDRDVLQGPWAKI 420
DB 425 EBEIRTNVPATEQVGSVSTNLQRNQAAATADVNTQGVLPGMVWQDRDVLQGPWAKI 484
QY 421 PHTDGHFHPSPMLMGFGGLKNPQIILIKNTPVPANPPAPFSAKTSFTQYSTQGVSE 480
DB 485 PHTDGHFHPSPMLMGFGGLKNPQIILIKNTPVPANPPAPFSAKTSFTQYSTQGVSE 544
QY 481 IEWELQKENSKEWNPVEQVTSNYAKSANVDFTVNNGLYTEPRPTGTRYLTRPL 534
DB 545 IEWELQKENSKEWNPVEQVTSNYAKSANVDFTVNNGLYTEPRPTGTRYLTRPL 598

RESULT 7
OS6652 PRELIMINARY; PRT; 735 AA.
AC OS6652;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Major coat protein VP1.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF043303; AAC03780.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein.
FT VARIANT 76 76 D -> V.
FT VARIANT 553 553 D -> N.
FT VARIANT 567 567 T -> N.
FT VARIANT 677 678 QV -> HV.
FT VARIANT 710 710 V -> R.
SQ SEQUENCE 735 AA; 81944 MW; 980BEEF46908390B CRC64;

Query Match 85.4%; Score 2481.5; DB 12; Length 735;
Best Local Similarity 83.3%; Pred. No. 3.3e-176;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGAPMADNNEGADGVGNAGSNWHCDSTWLGDRVITTTSTRTWALPTYNHLYKQISS 60
DB 203 MATGSGAPMADNNEGADGVGNAGSNWHCDSTWLGDRVITTTSTRTWALPTYNHLYKQISS 262
QY 61 ASTGASNDNHFGYSTPMGWYDFNRFCHFSPRDWRQLNNNMGPRPKRLNFKLFIQVK 120
DB 263 QS-GASNDNHFGYSTPMGWYDFNRFCHFSPRDWRQLNNNMGPRPKRLNFKLFIQVK 321
QY 121 EVTTNDGVTTIANNLTSTQVDSSEYQLPYVLGSAHQCLPFPADVFMIPOYGYLTIN 180
DB 322 EVTQNDGVTTIANNLTSTQVDSSEYQLPYVLGSAHQCLPFPADVFMIPOYGYLTIN 381

```

```
QY 181 NGSQAVGRSSFYCLEYFPPSOMLTGNNFTSFYFVPHSSYAHQSGLDRLMPLIDQY 240
Db 382 NGSQAVGRSSFYCLEYFPPSOMLTGNNFTSFYFVPHSSYAHQSGLDRLMPLIDQY 441
QY 241 LYYLNRTOQSGAQNKDLLFRSGSPAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNF 300
Db 442 LYYLNRTOQSGAQNKDLLFRSGSPAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNF 501
QY 301 WTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDNNMID 360
Db 502 WTGATKYHLNGRSLVNPFGVAMATHKDDERFPSSGVLI FGK--TGATNKTTLENVLM 561
QY 361 DEEIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 420
Db 562 DEEIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 621
QY 421 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 480
Db 622 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 681
QY 481 IEWELOKSKRNWPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534
Db 682 IEWELOKSKRNWPEIQYTSNYKSVNVDFTVDNGVYSEPRPIGTRYLTRNL 735

RESULT 8
QYQJGO PRELIMINARY; PRT; 737 AA.
AC Q8JQGO;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Capsid protein.
OS Adeno-associated virus 7.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=202812;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12192090;
RA Gao G.P., Alvira M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
RT human gene therapy.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859 (2002).
DR EMBL; AF513851; AA003855.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 737 AA; 81652 MW; D63B8CE7583140D0 CRC64;

Query Match 83.6%; Score 2430; DB 12; Length 737;
Best Local Similarity 82.1%; Pred. No. 2.3e-172;
Matches 440; Conservative 45; Mismatches 47; Indels 4; Gaps 3;

QY 1 MASGGAPMADNNEGADVGNAGNWHCDSTWLGDRVITSTRTWALPTNNHLYKQISS 60
Db 204 VAAGGAPMADNNEGADVGNAGNWHCDSTWLGDRVITSTRTWALPTNNHLYKQISS 263
QY 61 ASTGASNDNHFGYSTPFWGYDFNRFCHFSPRDQRLINNNWGPRLKPLFNIQVK 120
Db 264 ETAGSTNDNTYFGYSTPFWGYDFNRFCHFSPRDQRLINNNWGPRLKPLFNIQVK 323
QY 121 EVTTNDGVTTIANNLTSTQVPSDSEYQLPYVLGSAHQGCLPPFPADVFMIPOYGYLTN 180
Db 324 EVTTNDGVTTIANNLTSTQVPSDSEYQLPYVLGSAHQGCLPPFPADVFMIPOYGYLTN 383
QY 181 NGSQAVGRSSFYCLEYFPPSOMLTGNNFTSFYFVPHSSYAHQSGLDRLMPLIDQY 240
Db 382 NGSQAVGRSSFYCLEYFPPSOMLTGNNFTSFYFVPHSSYAHQSGLDRLMPLIDQY 441
QY 241 LYYLNRTOQSGAQNKDLLFRSGSPAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNF 299
Db 442 LYYLNRTOQSGAQNKDLLFRSGSPAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNF 503
QY 300 TWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDNNMID 359
Db 502 WTGATKYHLNGRSLVNPFGVAMATHKDDERFPSSGVLI FGK--TGATNKTTLENVLM 561
QY 361 DEEIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 419
Db 562 DEEIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 623
QY 420 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 479
Db 479 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 534
```

```
Db 504 AWTGATKYHLNGRSLVNPFGVAMATHKDDERFPSSGVLI FGK--TGATNKTTLENVLM 561
QY 359 TDEEIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 418
Db 562 TDEEIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 621
QY 419 KIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 478
Db 622 KIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 681
QY 479 VEIEWELOKSKRNWPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534
Db 682 VEIEWELOKSKRNWPEIQYTSNYKSVNVDFTVDNGVYSEPRPIGTRYLTRNL 737

RESULT 9
QYQJF8 PRELIMINARY; PRT; 738 AA.
AC Q8JQF8;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Capsid protein.
OS Adeno-associated virus 8.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=202813;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12192090;
RA Gao G.P., Alvira M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
RT human gene therapy.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859 (2002).
DR EMBL; AF513852; AA003857.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 738 AA; 81756 MW; 59BC2BE1B148AA14 CRC64;

Query Match 82.1%; Score 2384.5; DB 12; Length 738;
Best Local Similarity 80.2%; Pred. No. 5.6e-169;
Matches 429; Conservative 44; Mismatches 61; Indels 1; Gaps 1;

QY 1 MASGGAPMADNNEGADVGNAGNWHCDSTWLGDRVITSTRTWALPTNNHLYKQISS 60
Db 204 MASGGAPMADNNEGADVGNAGNWHCDSTWLGDRVITSTRTWALPTNNHLYKQISS 263
QY 61 -ASTGASNDNHFGYSTPFWGYDFNRFCHFSPRDQRLINNNWGPRLKPLFNIQVK 119
Db 264 GTSGATNDNTYFGYSTPFWGYDFNRFCHFSPRDQRLINNNWGPRLKPLFNIQVK 323
QY 120 KEVTTNDGVTTIANNLTSTQVPSDSEYQLPYVLGSAHQGCLPPFPADVFMIPOYGYLTN 179
Db 324 KEVTTNDGVTTIANNLTSTQVPSDSEYQLPYVLGSAHQGCLPPFPADVFMIPOYGYLTN 383
QY 180 NGSQAVGRSSFYCLEYFPPSOMLTGNNFTSFYFVPHSSYAHQSGLDRLMPLIDQY 239
Db 384 NGSQAVGRSSFYCLEYFPPSOMLTGNNFTSFYFVPHSSYAHQSGLDRLMPLIDQY 443
QY 240 LYYLNRTOQSGAQNKDLLFRSGSPAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNF 299
Db 444 LYYLNRTOQSGAQNKDLLFRSGSPAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNF 503
QY 300 TWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDNNMID 359
Db 504 AWTGATKYHLNGRSLVNPFGVAMATHKDDERFPSSGVLI FGK--TGATNKTTLENVLM 563
QY 360 DEEIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 419
Db 564 SEEBIKATNPVATERFQGTVAIVNFQSSSTD PATGDVHAMGALPGMWQDRDYLQGP IWA 623
QY 420 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 479
Db 479 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKPSAFITQYSTGVSV 534
```

```
Db 624 IPHTDGNFHPSPLMGGFGLKHPPQILIKNTVPADPTTFNQSKLNSFIQYSTGQVSV 683
Qy 480 ETEWELQKENSXKWNPEVOYTSNYAKSANVDFTVNNGLYTPRPIGTYRLTP 534
Db 684 ETEWELQKENSXKWNPEVOYTSNYAKSANVDFTVNNGLYTPRPIGTYRLTP 738

RESULT 10
Q67668 PRELIMINARY; PRT; 534 AA.
ID Q67668
AC Q67668
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE VP3.
GN VP3.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RL associated virus 2.";
RL Virology 212:562-573(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Virulent B;
RA Zadori Z.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25749; AAA83232.1; -.
DR InterPro; IPR001403; Parvo.coat.
DR Pfam; PF00740; Parvo.coat; 1.
SQ SEQUENCE 534 AA; 5996 MW; F1F049558EACE92 CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 534;
Best Local Similarity 57.1%; Pred. No. 4.5e-118;
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

Qy 1 MASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVITSTRTWALPTNNHLYKQISS 60
Db 1 MAEGGGAGMGDSGGADGVGNASGNWHCDSQMMGNTVITKTRTWLPSYNNHIYKAITS 60
Qy 61 ASTGASNDNH--YFGYSTPWGYFDNRFCHFSPRDWORLNNHNGFRPKLNPKLENIQ 118
Db 61 ---GTSQDANVQYAGYSTPWGYFDNRFCHFSPRDWORLNNHNGIRPKSLKFIENVQ 117
Qy 119 VKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQGLPPFPADVFMIPOYGYLT 178
Db 118 VKEVTTDQTKTIANNLTSTIQVTFDDEHQLPYVLGSATEGTMPPFPSSDVYALPOGYCT 177
Qy 179 LN---NGSQAVGRSSFYCYLEFPFSQMLRTGNNTFTSYTFEVPFHSVAHSQSLRLMNP 235
Db 178 MHTNQNGARFNDRSAFYCYLEFPFSQMLRTGNNTFTSYTFEVPFHSVAHSQSLRLMNP 237
Qy 236 LIDQYLYLNRNTQNSGSAQNKDLFSPGSPAGMSVQPKNWLPGPCYRQORV-SKTKTDN 294
Db 238 LVDQYLYNFEV-DSSRNAQ-----PKAVKAGYGTWGRNWLPGPKFLDQVRAYTGGTD 291
Qy 295 NNSNFT-WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSGVMIFGKE--SAGASNT 351
Db 292 NYANWNIWSNGKNVLDKQRYLLQPGVSATYTEGEASSLPQNILGIATKADPYRSGSTTA 351
Qy 352 ALDNYMITDEEIKATNPVATERFGTVAVNFQSSSDPATGTVHANGALPGMWQDRDVI 411
Db 352 GISDINMTVEQVAPNGVGWPKYGRVTNEQNTTAPTSSDLVLGALPGMWQNRDIY 411
Qy 412 LQGPICAKPKTDGKHPSPNGLGGFLHNPFPQVFIKNTVPADPVEYVHQWNSYITQ 471
Db 412 LQGPICAKPKTDGKHPSPNGLGGFLHNPFPQVFIKNTVPADPVEYVHQWNSYITQ 471
```

```
Qy 472 YSTGQSVSEIWELOKENSXKWNPEVOYTSNYAKSANVDFTVNNGLYTPRPIGTYRLT 531
Db 472 YSTGQSVSEIWELOKENSXKWNPEVOYTSNYAKSANVDFTVNNGLYTPRPIGTYRLT 531
Qy 532 RPL 534
Db 532 QNL 534

RESULT 11
Q67667 PRELIMINARY; PRT; 587 AA.
ID Q67667
AC Q67667
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE VP2.
GN VP2.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RL associated virus 2.";
RL Virology 212:562-573(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Virulent B;
RA Zadori Z.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25749; AAA83232.1; -.
DR InterPro; IPR001403; Parvo.coat.
DR Pfam; PF00740; Parvo.coat; 1.
SQ SEQUENCE 587 AA; 65246 MW; 0278B5D5FE7F0423 CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 587;
Best Local Similarity 57.1%; Pred. No. 5.2e-118;
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

Qy 1 MASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVITSTRTWALPTNNHLYKQISS 60
Db 54 MAEGGGAGMGDSGGADGVGNASGNWHCDSQMMGNTVITKTRTWLPSYNNHIYKAITS 113
Qy 61 ASTGASNDNH--YFGYSTPWGYFDNRFCHFSPRDWORLNNHNGFRPKLNPKLENIQ 118
Db 114 ---GTSQDANVQYAGYSTPWGYFDNRFCHFSPRDWORLNNHNGIRPKSLKFIENVQ 170
Qy 119 VKEVTNDGVTTIANNLTSTVQFSDSEYQLPYVLGSAHQGLPPFPADVFMIPOYGYLT 178
Db 171 VKEVTTDQTKTIANNLTSTIQVTFDDEHQLPYVLGSATEGTMPPFPSSDVYALPOGYCT 230
Qy 179 LN---NGSQAVGRSSFYCYLEFPFSQMLRTGNNTFTSYTFEVPFHSVAHSQSLRLMNP 235
Db 231 MHTNQNGARFNDRSAFYCYLEFPFSQMLRTGNNTFTSYTFEVPFHSVAHSQSLRLMNP 290
Qy 236 LIDQYLYLNRNTQNSGSAQNKDLFSPGSPAGMSVQPKNWLPGPCYRQORV-SKTKTDN 294
Db 291 LVDQYLYNFEV-DSSRNAQ-----PKAVKAGYGTWGRNWLPGPKFLDQVRAYTGGTD 344
Qy 295 NNSNFT-WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSGVMIFGKE--SAGASNT 351
Db 345 NYANWNIWSNGKNVLDKQRYLLQPGVSATYTEGEASSLPQNILGIATKADPYRSGSTTA 404
Qy 352 ALDNYMITDEEIKATNPVATERFGTVAVNFQSSSDPATGTVHANGALPGMWQDRDVI 411
Db 405 GISDINMTVEQVAPNGVGWPKYGRVTNEQNTTAPTSSDLVLGALPGMWQNRDIY 464
```



```

QY 412 LQGPWAKIPHTDGHFHPSPLMGGFGLKNPPQIILKNTVPANPPAEFSATKPFASFTQ 471
DB 465 LQGPWAKIPHTDGHFHPSPLMGGFGLKNPPQIILKNTVPANPPAEFSATKPFASFTQ 524
QY 472 YSTGQVSVELEWELOKENSKRMNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIGTRYLT 531
DB 525 YSTGQCTVMWELKENSCKRWPEIQFTSNFNRISIMFAPNETGGYVEDRLIGTRYLT 584
QY 532 RPL 534
DB 585 QNL 587

RESULT 12
Q67666 PRELIMINARY; PRT; 732 AA.
AC Q67666;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VPI.
GN VP.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Virulent B;
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Virology 212:562-573(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Virulent B;
RA Zadori Z.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25749; AAA83230.1; -.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat; 1.
SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 732;
Best Local Similarity 57.1%; Pred. No. 7.1e-118;
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

QY 1 MASGGAPWADNNEGADGVGNASGNHCHDSTWLGDRVITSTRTWALPTYNHLYKQISS 60
DB 199 MABGGGAGMDSGGADGVGNASGNHCHDSTWLGDRVITSTRTWALPTYNHLYKQISS 258
QY 61 ASTGASNDH--YFGYSTPWGYDFNRFCHFSPRDWORLNNWGRPKLNFKLFNIQ 118
DB 259 ---GTSQDANVQAGYSTPWGYDFNRFCHFSPRDWORLNNWGRPKLNFKLFNIQ 315
QY 119 VKEVTTNDGVTIANNLTSTVQFSDSEYQLPYVLGSAHQCLPPFPADVFMIPOGYLT 178
DB 316 VKEVTTDQTKTIANNLTSTIQVTDDEHQLPYVLGSATEGTMPPPSDVYALPOGYCT 375
QY 179 LN---NGSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSVAHQSGLDLRLNP 235
DB 376 MHTNQNGARFNDRAFICYCLEYFPSPQMLRTGNFTFEEDFEEVPHSMFAHQSGLDLRLNP 435
QY 236 LIDQYLYLNRTONOSGSAQNKDLLFSRGSAGMSVQPKNWLPGPCYQOQRY-SKTKDN 294
DB 436 LVDQYLYLNFNEV--DSSRNAQ-----FKKAVKAGYGTMGRLNLPGLKLDQVRATYTGTD 489
QY 295 NNSNFT-WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSGVMI FGKE--SAGASNT 351
DB 490 NYANWNIWSNGKNVNLKDRQYLLQPGVSAATHTKVEASSIPAQNILGLAKOPYRSGSTTA 549
QY 352 ALDNNWITDEEIEIKATNPATERFGTVAVNFQSSSTDPATGVDHANGALPGMWQDRDVI 411
DB 490 GISDIMVTDEGEVAPTNGVGWPKYKTKVTNEQNTTAPTSSDLVLGALPGMWQNRDIY 609
QY 412 LQGPWAKIPHTDGHFHPSPLMGGFGLKNPPQIILKNTVPANPPAEFSATKPFASFTQ 471

```

```

DB 550 GISDIMVTDEGEVAPTNGVGWPKYKTKVTNEQNTTAPTSSDLVLGALPGMWQNRDIY 609
QY 412 LQGPWAKIPHTDGHFHPSPLMGGFGLKNPPQIILKNTVPANPPAEFSATKPFASFTQ 471
DB 610 LQGPWAKIPHTDGHFHPSPLMGGFGLKNPPQIILKNTVPANPPAEFSATKPFASFTQ 669
QY 472 YSTGQVSVELEWELOKENSKRMNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIGTRYLT 531
DB 670 YSTGQCTVMWELKENSCKRWPEIQFTSNFNRISIMFAPNETGGYVEDRLIGTRYLT 729
QY 532 RPL 534
DB 730 QNL 732

RESULT 13
Q8V395 PRELIMINARY; PRT; 732 AA.
AC Q8V395;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Capsid protein VP.
GN VP.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPV-YG;
RA Ge Y., You Y., Xu Q.;
RT "Analysis of the major open reading frames' nucleotide sequences in
RT Goose parvovirus GPV-YG strain isolated in China.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF416726; AAL37722.1; -.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat; 1.
SQ SEQUENCE 732 AA; 81456 MW; 73F2E48C769744B6 CRC64;

Query Match 58.4%; Score 1697; DB 12; Length 732;
Best Local Similarity 57.3%; Pred. No. 8.4e-118;
Matches 311; Conservative 82; Mismatches 132; Indels 18; Gaps 8;

QY 1 MASGGAPWADNNEGADGVGNASGNHCHDSTWLGDRVITSTRTWALPTYNHLYKQISS 60
DB 199 MABGGGAGMDSGGADGVGNASGNHCHDSTWLGDRVITSTRTWALPTYNHLYKQISS 258
QY 61 ASTGASNDH--HYFGYSTPWGYDFNRFCHFSPRDWORLNNWGRPKLNFKLFNIQ 118
DB 259 ---GTSQDANVQAGYSTPWGYDFNRFCHFSPRDWORLNNWGRPKLNFKLFNIQ 315
QY 119 VKEVTTNDGVTIANNLTSTVQFSDSEYQLPYVLGSAHQCLPPFPADVFMIPOGYLT 178
DB 316 VKEVTTDQTKTIANNLTSTIQVTDDEHQLPYVLGSATEGTMPPPSDVYALPOGYCT 375
QY 179 LN---NGSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSVAHQSGLDLRLNP 235
DB 376 MHTNQNGARFNDRAFICYCLEYFPSPQMLRTGNFTFEEDFEEVPHSMFAHQSGLDLRLNP 435
QY 236 LIDQYLYLNRTONOSGSAQNKDLLFSRGSAGMSVQPKNWLPGPCYQOQRY-SKTKDN 294
DB 436 LVDQYLYLNFNEV--DSSRNAQ-----FKKAVKAGYGTMGRLNLPGLKLDQVRATYTGTD 489
QY 295 NNSNFT-WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSGVMI FGKE--SAGASNT 351
DB 490 NYANWNIWSNGKNVNLKDRQYLLQPGVSAATHTKVEASSIPAQNILGLAKOPYRSGSTTA 549
QY 352 ALDNNWITDEEIEIKATNPATERFGTVAVNFQSSSTDPATGVDHANGALPGMWQDRDVI 411
DB 550 GISDIMVTDEGEVAPTNGVGWPKYKTKVTNEQNTTAPTSSDLVLGALPGMWQNRDIY 609
QY 412 LQGPWAKIPHTDGHFHPSPLMGGFGLKNPPQIILKNTVPANPPAEFSATKPFASFTQ 471

```



Db	612	GPIWAKIPKTDGKFHSPNLGGFGLHNPFPQVFIKNTFVPADPPLEYNQKWSYITQYS	671
Qy	474	TGOVSVEIEWELQKENSKEWNPVOYTSNYAKSANYDETVDNNGLYTERPPIGTRYLTTP	533
Db	672	TGQCTVEMWELRKENSKEWNPFIQFTSNFGNRTSTMFAPNETGGYVEDRLIGTRYLTQN	731
Qy	534	L 534	
Db	732	L 732	

Search completed: January 21, 2004, 16:06:11  
Job time : 28.7143 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:01:27 ; Search time 12.2857 Seconds  
(without alignments)  
1839.046 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGGAPMADNEGADGVG.....NNGLYTEPRPIGTRYLTRPL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata1/iaa/5A COMB.pep:\*
- 2: /cgn2\_6/prodata1/iaa/5B COMB.pep:\*
- 3: /cgn2\_6/prodata1/iaa/6A COMB.pep:\*
- 4: /cgn2\_6/prodata1/iaa/6B COMB.pep:\*
- 5: /cgn2\_6/prodata1/iaa/PCUS COMB.pep:\*
- 6: /cgn2\_6/prodata1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2481.5	85.4	735	4	US-09-321-589-1
2	1690.5	58.2	544	4	US-09-532-594B-18
3	1690.5	58.2	598	4	US-09-532-594B-16
4	1690.5	58.2	734	4	US-09-532-594B-4
5	491.5	16.9	756	4	US-09-438-268-4
6	479.5	16.5	543	3	US-08-856-841-22
7	326	11.2	415	3	US-08-856-841-20
8	318	10.9	395	3	US-08-856-841-13
9	318	10.9	486	3	US-08-856-841-19
10	317.5	10.9	500	3	US-08-856-841-16
11	317.5	10.9	501	3	US-08-856-841-18
12	304.5	10.5	264	3	US-08-856-841-14
13	303	10.4	398	3	US-08-856-841-21
14	209.5	7.2	579	6	5223424-13
15	183	6.3	584	3	US-09-022-949-2
16	171	5.9	387	3	US-08-856-841-17
17	119	4.1	1095	4	US-09-107-532A-3855
18	111	3.8	655	1	US-08-469-202-27
19	111	3.8	655	2	US-08-484-434C-34
20	111	3.8	655	4	US-09-384-361-34
21	109	3.8	1186	1	US-08-485-568A-4
22	109	3.8	1186	2	US-08-357-698-6
23	109	3.8	1186	2	US-08-590-554A-4
24	109	3.8	1186	2	US-09-184-223-4
25	109	3.8	1186	5	PCT-US93-12682-6
26	108	3.7	624	3	US-08-947-965-78
27	108	3.7	655	1	US-08-469-202-28

28	108	3.7	655	2	US-08-484-434C-35	Sequence 35, Appl
29	108	3.7	655	4	US-09-384-361-35	Sequence 35, Appl
30	104.5	3.6	717	4	US-09-626-589-1	Sequence 1, Appl
31	104	3.6	1013	3	US-09-415-522-8	Sequence 8, Appl
32	102	3.5	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
33	99.5	3.4	2736	4	US-09-252-991A-30227	Sequence 30227, A
34	99	3.4	824	4	US-09-626-589-3	Sequence 3, Appl
35	98.5	3.4	997	1	US-08-233-540-1	Sequence 1, Appl
36	98.5	3.4	997	1	US-08-428-949A-1	Sequence 1, Appl
37	98.5	3.4	997	1	US-08-428-948A-1	Sequence 1, Appl
38	98.5	3.4	997	2	US-08-428-946-1	Sequence 1, Appl
39	98.5	3.4	997	5	PCT-US95-04656-1	Sequence 1, Appl
40	98.5	3.4	998	1	US-08-233-008A-6	Sequence 6, Appl
41	98.5	3.4	1000	4	US-09-193-562D-30	Sequence 30, Appl
42	98.5	3.4	1021	1	US-08-233-008A-2	Sequence 2, Appl
43	98	3.4	856	4	US-09-328-352-7255	Sequence 7255, Ap
44	98	3.4	1178	6	5254799-5	Patent No. 5254799
45	97	3.3	540	3	US-08-687-580B-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1  
US-09-321-589-1  
; Sequence 1, Application US/09321589  
; Patent No. 6498244  
; GENERAL INFORMATION:  
; APPLICANT: PATEL, SALIL D.  
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS  
; FILE REFERENCE: 39572  
; CURRENT APPLICATION NUMBER: US/09/321,589  
; CURRENT FILING DATE: 1999-05-28  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: Adeno-associated virus  
US-09-321-589-1

Query Match 85.4%; Score 2481.5; DB 4; Length 735;  
Best Local Similarity 83.3%; Pred. No. 1.6e-218;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

Qy	1	MASGGGAPMADNEGADGVGNAGNWHCDSTWLGDRVITTTTRTWTALPTYNHLYKQISS	60
Db	203	MATSGGAPMADNEGADGVGNAGNWHCDSTWLGDRVITTTTRTWTALPTYNHLYKQISS	262
Qy	61	ASTGASNDNHFGYSTPMGYDFNRFCHFSFRDQWRLNNNNGFRPKRLNFKLFNIQVK	120
Db	263	QS-GASNDNHFGYSTPMGYDFNRFCHFSFRDQWRLNNNNGFRPKRLNFKLFNIQVK	321
Qy	121	EVTNDGVTTIANNLTSTVQFSDSEYQLPYVLSAHQGLPPPPADVFMIPOYGYLTIN	180
Db	322	EVTQNDGVTTIANNLTSTVQFSDSEYQLPYVLSAHQGLPPPPADVFMIPOYGYLTIN	381
Qy	181	NGSQAVGRSSPYCYLEYPPSQMLRTGNFTFSTTEVPFHSYAHQSQSLDRMLNLIQY	240
Db	382	NGSQAVGRSSPYCYLEYPPSQMLRTGNFTFSTTEVPFHSYAHQSQSLDRMLNLIQY	441
Qy	241	LYLNRTOQSGSAQNKDLFSGSPAGMSVQPKNWLPGCVROORVSKTKTDNNNSNPT	300
Db	442	LYLSRTNTPSGTTTQSLQFQAGASDIRQSRNWLPGPCYRQORVSKTSADNNSEYS	501
Qy	301	WTGASKYNLNGRESIINPTAMASHKODEDEKFPMSGVMIFGKSAGASNTALDNVMTD	360
Db	502	WTGATKYHLNGRDSLNVNPGPAMASHKODEDEKFPMSGVMIFGKSAGASNTALDNVMTD	561
Qy	361	EEETKATNPVATERPGTVAVNPFQSSSTDPATGDYHMGALPGMWQDRDVLQGPWAKI	420
Db	562	EEETKATNPVATERPGTVAVNPFQSSSTDPATGDYHMGALPGMWQDRDVLQGPWAKI	621

QY 421 PHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPEASITQYSTGVSE 480  
DB 622 PHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPEASITQYSTGVSE 681  
QY 481 IEWELQKSKRNPEVQVTSYAKSANVDFTVDDNGLYTEPRPIGTLYLTP 534  
DB 682 IEWELQKSKRNPEVQVTSYAKSANVDFTVDDNGLYTEPRPIGTLYLTP 735

## RESULT 2

US-09-532-594B-18  
; Sequence 18, Application US/09532594B  
; Patent No. 6468524  
; GENERAL INFORMATION:  
; APPLICANT: Chorini, John A.  
; APPLICANT: Kotin, Robert M.  
; APPLICANT: Safer, Brian  
; APPLICANT: Davidson, Beverly  
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF  
; FILE REFERENCE: 14014.025202  
; CURRENT APPLICATION NUMBER: US/09/532,594B  
; CURRENT FILING DATE: 2000-03-22  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 544  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =  
; OTHER INFORMATION: synthetic construct  
; NAME/KEY: misc feature  
; OTHER INFORMATION: AAV4 capsid protein VP3  
US-09-532-594B-18

Query Match 58.2%; Score 1690.5; DB 4; Length 544;  
Best Local Similarity 59.4%; Pred. No. 2.8e-146;  
Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;  
QY 2 ASGGGAPMADNNEGADGCVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYKQISSA 61  
DB 9 AAAGGA-EGGQAGDVGNGASGDHWCSTWSEGHVTTTSTRTWVLPYNNHLYKRLGE- 66  
QY 62 STGASNDNHYFGYSTPWGDFDNRHCHFSRDMORLINNNWGFPRKLNKLFNIQVKE 121  
DB 67 ---SLOSTNTYNGFSTPWGDFDNRHCHFSRDMORLINNNWGFPRKLNKLFNIQVKE 123  
QY 122 VTNDGVTIANNLTSTVQVPSDSEYQLPYVLGSAHQGLPPFPADVFMIPQYGV---LT 178  
DB 124 VTTSGETTVAANNLTSTVQIFADSSYELPYVMDAGQESLPPFPNDVFMVQYGCGLVT 183  
QY 179 LNNGSAVGRSSFYCLEYFPQMLRTGNFTFSYTFEBVPHSSYAHSSQSLDRLMPLID 238  
DB 184 GNTSQOQTDRAFYCLEYFPQMLRTGNFTFSYTFEBVPHSSYAHSSQSLDRLMPLID 243  
QY 239 QYLYLNRNTQN---QSGSAQNKDLLFSRGSAGMSVQPKWLPGCYRQQRVSKTKDN 294  
DB 244 QYLMGLQSTTTGTLNAGTATTN---FTKLRPTNFSNFKQWLPGSIKQOGFSKTA--N 298  
QY 295 NNSNFTWTGAS---KY-----NLNGRESIINPGTAMASHKDDKDFPFMSGVMI FGKESAG 347  
DB 299 QNYKIPATGSDSLIKYETHSTLDGRWSALTTPGPMATAGPADSK--PSNQLIFAGPKQNG 357  
QY 348 ASNTALDNVMTDEBEIKATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMVWD 407  
DB 358 NTATVPGTLIFTSEBELAATNATDMDWGNLPGDQSNLPTVDRLTALGAVPGMVWQN 417  
QY 408 RDVYLGQPIWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPA 467  
DB 418 RDIYQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPA 477  
QY 468 FITQYSTGVSEVIEWELQKSKRNPEVQVTSYAKSANVDFTVDDNGLYTEPRPIGT 527

DB 478 FITQYSTGVSEVIEWELQKSKRNPEVQVTSYAKSANVDFTVDDNGLYTEPRPIGT 537  
QY 528 RYLTRPL 534  
DB 538 RYLTHHL 544

## RESULT 3

US-09-532-594B-16  
; Sequence 16, Application US/09532594B  
; Patent No. 6468524  
; GENERAL INFORMATION:  
; APPLICANT: Chorini, John A.  
; APPLICANT: Kotin, Robert M.  
; APPLICANT: Safer, Brian  
; APPLICANT: Davidson, Beverly  
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF  
; FILE REFERENCE: 14014.025202  
; CURRENT APPLICATION NUMBER: US/09/532,594B  
; CURRENT FILING DATE: 2000-03-22  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 598  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =  
; OTHER INFORMATION: synthetic construct  
; NAME/KEY: misc feature  
; OTHER INFORMATION: AAV4 capsid protein VP2  
US-09-532-594B-16

Query Match 58.2%; Score 1690.5; DB 4; Length 598;  
Best Local Similarity 59.4%; Pred. No. 3.2e-146;  
Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;  
QY 2 ASGGGAPMADNNEGADGCVGNAGNWHCDSTWLGDRVITTTSTRTWALPTNNHLYKQISSA 61  
DB 63 AAAGGA-EGGQAGDVGNGASGDHWCSTWSEGHVTTTSTRTWVLPYNNHLYKRLGE- 120  
QY 62 STGASNDNHYFGYSTPWGDFDNRHCHFSRDMORLINNNWGFPRKLNKLFNIQVKE 121  
DB 121 ---SLOSTNTYNGFSTPWGDFDNRHCHFSRDMORLINNNWGFPRKLNKLFNIQVKE 177  
QY 122 VTNDGVTIANNLTSTVQVPSDSEYQLPYVLGSAHQGLPPFPADVFMIPQYGV---LT 178  
DB 178 VTTSGETTVAANNLTSTVQIFADSSYELPYVMDAGQESLPPFPNDVFMVQYGCGLVT 237  
QY 179 LNNGSAVGRSSFYCLEYFPQMLRTGNFTFSYTFEBVPHSSYAHSSQSLDRLMPLID 238  
DB 238 GNTSQOQTDRAFYCLEYFPQMLRTGNFTFSYTFEBVPHSSYAHSSQSLDRLMPLID 297  
QY 239 QYLYLNRNTQN---QSGSAQNKDLLFSRGSAGMSVQPKWLPGCYRQQRVSKTKDN 294  
DB 298 QYLMGLQSTTTGTLNAGTATTN---FTKLRPTNFSNFKQWLPGSIKQOGFSKTA--N 352  
QY 295 NNSNFTWTGAS---KY-----NLNGRESIINPGTAMASHKDDKDFPFMSGVMI FGKESAG 347  
DB 353 QNYKIPATGSDSLIKYETHSTLDGRWSALTTPGPMATAGPADSK--FNSQLIFAGPKQNG 411  
QY 348 ASNTALDNVMTDEBEIKATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMVWD 407  
DB 412 NTATVPGTLIFTSEBELAATNATDMDWGNLPGDQSNLPTVDRLTALGAVPGMVWQN 471  
QY 408 RDVYLGQPIWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPA 467  
DB 472 RDIYQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTVPANPPAEFSATKPA 531  
QY 468 FITQYSTGVSEVIEWELQKSKRNPEVQVTSYAKSANVDFTVDDNGLYTEPRPIGT 527  
DB 532 FITQYSTGVSEVIEWELQKSKRNPEVQVTSYAKSANVDFTVDDNGLYTEPRPIGT 591



Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 543  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:

PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 22:  
US-08-856-841-22  
Query Match 16.5%; Score 479.5; DB 3; Length 543;  
Best Local Similarity 27.1%; Pred. No. 2.3e-35;  
Matches 141; Conservative 79; Mismatches 226; Indels 75; Gaps 16;  
QY 10 ADNNEGADGVGNAS--GNWHCDSTWLGDRVITTTSTWALPTYNHLYKQISSASTGASN 67  
Db 16 AEAAGTGGGGNSVSKMSEGAIFSANSTCFSSQFLIPYDEHYHYKVFSPAASCHN 75  
QY 68 D-----NHYFGYSTPWGYFDNFRHCHSPRDWQRLNNNMGRKRLNFKLFI 117  
Db 76 ASGKEAKVCTISPIMGYSTPWYLDNFALNLFSPLEFQHLIENYGSIAADALTVTITSEI 135  
QY 118 QVKEVT--TNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHQCLPPFPADYPMIQYG 175  
Db 136 AVKDVTDTKGGV--QVTDSTTGLCLMLVDHEYKYPYVLGGQDTLAPELFIWVTFPPQYA 194  
QY 176 YLTLLN--GSAQVG-----RSSFYCLEYFPSPQMLRTGNNTFSYTFEEVFPFSSYAH 225  
Db 195 YLTGVDTNTOIGSDSKLASBSAFYVLEHSSFOLLGTGTASTMSYKFPVPENLEGC 254  
QY 226 SQSLDRMLNPLIDQYLYLNRNQSGSAQNKLFSRGSFAGMSVQPKWLPQPCVTRQQ 285  
Db 255 SQHFYEMYNPL---YGSRLGVPTDTLGGDPKFRSL-----THEDHAIQPNFMFGPLVNSV 306  
QY 286 RVSKTKDNNNSNFTWTGASKYNLNGRESIINPG--TAMASHKDDKEDKFFPMSCGMIFGKE 344  
Db 307 STKEGSSNTGAGKALTGLSTGTSQNTIRSLRPGVSPYHHWDTDKYVTGINAISHGQT 366  
QY 345 SAGASNTALDNV-----MITDEEIKATNPVATERFGTVAVNFQSSSTDPAITGDVHAM 397  
Db 367 TYG---NAEDKEYQQGVGRFNEKEQLKQLGLNMHTY-----FPNKGTOQYTDQIE-R 416  
QY 398 GALPGVMVQDRDVLQGPWAKPIHPTDGHFHS--PLMGFGFLKNPPQIILKNTPTVPANP 456  
Db 417 PLMVGSVMNRRLHYESQLMSKIPNLDDSPKTOFAALGGLHGLHQPPI----- 465  
QY 457 PAEFSAKTFASFITOYSTGOVSVEIEWEL--QKENSKEWNPE 496  
Db 466 -----FLKQYAVGIMVTMTFPLGPRKATGRWNFO 495  
RESULT 7  
US-08-856-841-20  
Sequence 20, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658





IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: I  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
US-08-856-841-13  
Query Match 10.9%; Score 318; DB 3; Length 395;  
Best Local Similarity 25.4%; Pred. No. 8.8e-21;  
Matches 99; Conservative 61; Mismatches 167; Indels 62; Gaps 13;  
QY 128 VTTIANLTSTQVFSDSYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTINN-GSOAV 186  
DB 1 MTWITPSLHACMLV--DHEIKYPIYVLGQDQTLAPLPIWVYFPQYATLVGDVNTQGI 58  
QY 187 G-----RSSFYCLEYFPQMLRTGNFTSYTFEEVPHSHYAHSSQSLDRMLNPLI 237  
DB 59 SGDSKKLASESAFYVLEHSSFQLLTGCTASMSYKFPVPPEENLEGCQSHPYMYNPL- 117  
QY 238 DOYLYLNTQNSGAQKDLFFSRGSPAGSVQPKWLPQPCYRQQRVSKTKTDNNNS 297  
DB 118 --YGSRLGVPDTLGGDPKFRSL-----THEDHAIQPNFMPGLVNSVSTKEGDSNTGA 170  
QY 298 NFTWTGSKYVNLNGRESIINPG-TAMASHKDDKDFPMSGVMIKESAGASNTALDNV 356  
DB 171 GKALTGLTGTQNTFRISLRPQVSPQYHHWTDKVTGAINAISHGQITYG---NABKE 227  
QY 357 -----MITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMYWQDRD 409  
DB 228 YQGVGRFPNEKEQLKQLGLNHHY-----FPNKGTOQYTDQIE-RLPMVGVSNRRA 280  
QY 410 VYLOGPIWAKIPTHGHPHS-PLMGFGGLKNPPQIILIKTTPVNPANPPAEFSATKFSF 468  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: I  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
US-08-856-841-19  
; Sequence 19, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 486  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE:  
; DESCRIPTION: PEPTIDE  
; HYPOTHEICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
; POSITION IN GENOME: N/A  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.

AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE: 1  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 19:  
US-08-856-841-19

Query Match 10.9%; Score 318; DB 3; Length 486;  
Best Local Similarity 32.8%; Pred.No. 1.2e-20;  
Matches 82; Conservative 31; Mismatches 109; Indels 28; Gaps 6;

QY 14 EGADGVGNASGN-----WICDSTWLGDRVITSTRTWALPTNNHLYKQISSASTCASND 68  
DB 235 EASTGAGGGSNPVKSMWSEGATFSANSVTCTFSRQFLIPYDPEHHYKVFSPAASCHNA 294  
QY 69 -----NHVFGYSTWGFDFNRHCHFSRDMQRLNNMNGFRPKRLNFKLFTIQ 118  
DB 295 SGKEAKVCTISPIMGYSTPWRVLDNALNLFSPLEFQHLIENYGSTAPDALTVTISETA 354  
QY 119 VKEVT--TNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQCLPPFPADVFMIPQCY 176  
DB 355 VKDVTDTGGV-QVTDSATGRCLMLVDHEYKYVYVLGGQDTLAPELFIWYFPPQVAY 413  
QY 177 LTINN-GSOAVG-----RSSCYCLEYFPQSQMLRTGNFTFSYTFEEVPFHSSYAH 226  
DB 414 LTVDVNTQGISGDSKKLASBESAFVLEHSSFQLLGTGTATMSYKFPVPPENLEGS 473  
QY 227 QSLDRMLNPL 236  
DB 474 QHFYMYNPL 483

RESULT 10  
US-08-856-841-16  
Sequence 16, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESS: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991

APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 16:  
PUBLICATION INFORMATION:  
AUTHORS: MANTATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES:  
DATE: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 16:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME: 67  
ISSUE:  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:

US-08-856-841-16

	Query Match	10.9%	Score 317.5;	DB 3;	Length 500;
	Best Local Similarity	33.1%	Pred. No. 1.4e-20;		
	Matches 83; Conservative 34;	Mismatches 109;	Indels 25;	Gaps 6;	
QY	10 ADNNSGADGVGNAS--GNWHCDSTWLGDRVITTTSTRTWTALPTYNHLYKQISSASTGSIN	67	:	:	:
Dd	242 AEASTGAGGGSGSNSVKMSMWSEGAFTSANSVTCTFSRQLFLIPDPEHHYKVFPSPAASHCN	301	:	:	:
QY	68 D-----NHYPGISTPWGYEDFNRFCHSPRDMDQLINNKGPPFKLNFKNFI	117	:	:	:
Dd	302 ASGKEAKVCTTSPINGYSPTWRYLDFNALMFFSPLEFQHLIENYGSIAPDALTTVTISEI	361	:	:	:
QY	118 QVKEVT---TNDSGVTTIANLLTSVVQSFDSEYQLPVVLGSAHGCLPFPDPADVFMIPQG	175	:	:	:
Dd	362 AVKDVTDKTGCV-QVTDSTTGRICMLVDHEYKYVPVVGQQDTLAPELPVWVFPPQVA	420	:	:	:
QY	176 YLTLAN-GSQAVG-----RSSFCLEYFPSQMLRTGNNTFTSVTFEETVPHSSYAH	225	:	:	:
Dd	421 YUTGDVNVTQIGSDSKKLASEERAFVTLHSFPQLTGCTGTASMKYKPFPVPPDENLEG	480	:	:	:
QY	226 SQSLDLRLMNPL	236	:	:	:
Dd	481 SQHFYEYMINPL	491	:	:	:

RESULT 11  
US-08-856-841-18  
Sequence 18, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:

```

LENGTH: 501
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.B.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: I
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 18:
US-08-856-841-18

Query Match 10.9%; Score 317.5; DB 3; Length 501
Best Local Similarity 33.1%; Pred. No. 1.4e-20;
Matches 83; Conservative 34; Mismatches 109; Indels 25

QY 10 ADNNEGADGVGNAS--GNWHCDSTWLGDRVITTTTRTWALPTNNHLYKQIGI
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 243 AEASTGAGGGGSNSVKSMWSEGATSANSVTCTPSRQFLIPYDPEHYKVP
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 D-----NHYGYSTPWGYFDNRFPHCHFSPRDMQRLNNNWGFRPFRK
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 303 ASGKEAKVCTISPIMGYSTPWRYLDFNALNLFPSPLEFQHLIENTYGSIA
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 QVKEVT--TNDGVTTIANLSTVOVFSDSVQLPYVLGSAHQGCLPPPPA
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 363 AKVDVDTKGGGV-QVTDSTGRLCMLVDHETKYPYVLQGGQDTLAPELPY
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 YLTLLN-GSQAVG-----RSSFYCLEYPPSQMLRTGNNTFFSYTYPEE
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 422 YLTGVDVNTQGISGDSKKLASEESAFVLEHSSFOLLGTGTASMSYKFP
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 226 SQSLDLNANPL 236
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 482 SQHFYEMTNPL 492
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-08-856-841-14
Sequence 14, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE

```

```

RESULT 12
US-08-856-841-14
; Sequence 14, Application US/08856841
; Patent No. 6274307
; GENERAL INFORMATION:
; APPLICANT: ERWIN SOUTSCHEK
; APPLICANT: MANFRED MOTZ
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROOKS HAYDT HAPFNER & DELAHUNTY
; STREET: 99 PARK AVENUE
;

```

CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 264  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PEPTIDE  
DESCRIPTION: N/A  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURES:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: I  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:

PAGES:  
DATE: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
US-08-856-841-14  
Query Match 10.5%; Score 304.5; DB 3; Length 264;  
Best Local Similarity 32.5%; Pred. No. 8.2e-20;  
Matches 76; Conservative 30; Mismatches 105; Indels 23; Gaps 5;  
QY 7 APWADNNEGADGVGNASGNWHCHDSTWLGDRVITTTSTRTWALPTYNHLYKQISSASTGAS 66  
DB 9 AAESTGAGGGGNSVKSMWSEGAFTSANSVCTFSRQFLIPYDPEHHYKVFSPAASCH 68  
QY 67 ND-----NHYFGYSTPMGYDFNRFHCHPSRDMQRLNNNNWGFPRKRLNPKLFN 116  
DB 69 NASGKEAKVCTISPIMGYSTPMRYLDFNALNLFPSPLSEFQHLIENYGSTAPDALTVTISE 128  
QY 117 IOVKEVT--TNDGVTTIANLLTSTVQFSDSEYQIPLYVLSAHQOCLPPFPADVFMIPQY 174  
DB 129 IAKDVTDKTKGGV-QVTDSTTGRLCMLVDHBYKYPIYVLGGQDTLAPELPIWVYFPQY 187  
QY 175 GYLTLNN-GSQAVG-----RSSFCYCLEVFPQMLRTGNNTFSYTFEBVP 218  
DB 188 AYLTVGDVNTQISGDSKKLASESAFTVLEHSSFQLLGCTGTASMSYKFPVP 241  
RESULT 13  
US-08-856-841-21  
Sequence 21, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994

```

; APPLICATION NUMBER: US 07/917,096
; FILING DATE: 4-AUGUST-1992
; APPLICATION NUMBER: PCT/DE91/00106
; FILING DATE: 8-FEBRUARY-1991
; APPLICATION NUMBER: DE40038262
; FILING DATE: 8-FEBRUARY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINSON, WILLIAM R.
; REGISTRATION NUMBER: 27,224
; REFERENCE/DOCKET NUMBER: LKR-9222-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-3355
; TELEFAX: (212) 557-5635
; TELEX: NONE
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: PEPTIDE
; HYPOTHETICAL: N/A
; ANTI-SENSE: N/A
; FRAGMENT TYPE: INTERNAL
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
; POSITION IN GENOME: N/A
; FEATURE:
; NAME/KEY: N/A
; LOCATION: N/A
; IDENTIFICATION METHOD: amino acid analysis and
; IDENTIFICATION METHOD: mass spectrometry
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: COSSART, Y.E.
; AUTHORS: FIELD, A.M.
; AUTHORS: CANT, B.
; AUTHORS: WIDDOWS, D.
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
; JOURNAL: LANCET
; VOLUME: I
; ISSUE:
; PAGES: 72 - 73
; DATE: 1975
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: 21:
; US-08-856-841-21

Query Match
Best Local Similarity 10.4%; Score 303; DB 3; Length 398;
Matches 96; Conservative 48; Mismatches 168; Indels 66; Gaps 12;

Qy 144 DSEYQLPYVLGSAHQCLPPFPADVMIPQYGYLTNN-GSOAVG-----RSSFYC 193
Db 4 DHEYKYPYVLGQODTLAPLEPIWYVPQYAYLTVDVNTQISGDSKCLASESAFYV 63
Qy 194 LEYFPQMLRTGNFTFSYTFEEVPHSSYAHQSGLRLMPLIDQYLYLNRTONQSGS 253
Db 64 LEHSPQLLGTGTATMSYKFPVPVENLEGCSQHFYEMYNPL---YGSRLGVPDTLGD 120
Qy 254 AQNKLLFSGRGPAGMSVQPKNLPGPCYQQRVSKTKTDNNNSFTWTGASKYLNNGRE 313
Db 121 PKFRLS-----THEDHAIQPMFPGPLVNSVSTKEGDSNTGAGKALTGLSTGTSQNT 175
Qy 314 SIINPG-TAMASHKDDKFFPMGSGVMIFGKESAGASNTALDNV-----MITDEEEK 365
Db 176 ISLRPGVSPQYHHWTDKYVTGAINAISHGQTYG---NAEDKEYQOQGVGRFPNKEBQLK 232
Qy 366 ATNPVATERFGTVANVFQSSSDPATGDVHAMGALPGMWNQDRDYLQPIWAKIPHTDG 425
Db 233 QLQGLNMHTY-----FPNKGTOQYTDQIE-RPLMVGSVNNRRALHYESQLMSKIPNLDD 285

; APPLICATION NUMBER: US/07/225,032
; FILING DATE: 27-JUL-1988
; APPLICATION NUMBER: 78,519
; FILING DATE: 27-JUL-1987
; APPLICATION NUMBER: 933,107
; FILING DATE: 20-NOV-1986
; APPLICATION NUMBER: 902,887
; FILING DATE: 02-SEP-1986
; APPLICATION NUMBER: 887,140
; FILING DATE: 17-JUL-1986
; APPLICATION NUMBER: 823,102
; FILING DATE: 27-JAN-1986
; APPLICATION NUMBER: 773,430
; FILING DATE: 06-SEP-1985
; SEQ ID NO:13:
; LENGTH: 579
5223424-13

Query Match
Best Local Similarity 7.2%; Score 209.5; DB 6; Length 579;
Matches 138; Conservative 80; Mismatches 237; Indels 153; Gaps 32;

Qy 3 SGGCAPMADNNEGADGVGNAGSNWH--CDSTWLGD---RVITSTRTWTALPYNNHLYKQ 57
Db 24 SGGGGG-GGGRGAGGVGVSTGSPFNQTFQYLGELVRITAHASRLIHLNMPHEHYTKR 82
Qy 58 IS--SASTGAS-----NDNHFGYSTPMGYDFDPNPFCHFPSPRDWORLNNNWGFRPKELN 111
Db 83 IHLNSESAGVAGQVODDAHTQMTVPMSLIDRNAWGVNFPADWQLISNNMTINLVSFE 142
Qy 112 FKLNIQVKEVT---TNDGVTTIANNLSTVQVPSDSEYQLPYVLGSAHQCLPPFPADV 168
Db 143 QEINNVVLKTTESATSPSKIYNDNLITASLWALDTNNTLPTTPAARSETLGFYFWLP 202
Qy 169 FMIPQYGYL-----TLNNGSQAVGRS-----SFYCLE-YFPQMLRTGNF 208
Db 203 TKPTQRYLLSCIRNLNLPPTVTGQSQITDSIQTGLSHDIMFYTIENAVPIHLRTGDEP 262
Qy 209 -TFSTTEEVPHSSYAHQSGLRLMPLIDQYLYLNRTONQSGSAGKDLLLSRGSQA 267
Db 263 STGYHFDTKPL--KLTHSMQTRN-----SLGLPP 290
Qy 268 GMSVQP-----KNWLPGPCYQ---QRVSKTKTD-----NNNSNFTWTGAS 305
Db 291 KVLTEPTTEGQHPGTLPGANTRKGYHQTINNSTEATARPQGVNTPTMNFDSYNGG 350
Qy 306 KYNLNGRESIINP--GTAMASHKDDKFFPMGSGVMIFGKESAGASNTALDNVITDDEE 363
Db 351 PF-----LTPIVPTADTQYDDE---PNGAIRPTMGYQHGLTTS-----SQBLE 392
Qy 364 IKATNP-----VATERFGTVV-NFQSSS-----TDPATG--DVHAMGALP-----GM 403

```

Db 393 RYTFNPSKCGRAPKQOFNOQAFLNLENTNGTLLPSDPIGGKSNKHFMTLNTYGLTA 452  
QY 404 VQDRDYLQGPWAKIPHTD--GHEHPSPLMGGEGLK-NPPQILIKNTVPVAPAEF 460  
Db 453 LNTAPVFNQGLWDEKLDLXPRH--VTAFVCKNPPQGLFKIAP--NLDDF 506  
QY 461 SA--TKFASITQSGQSVSEIEMELQKSNKRWNPVEQVTSYAKSANVDFVDNGL 518  
Db 507 NADSPQOPRIITD-SNFWMGKTLTFTAKMRSSNMNPIQHTT-----TAENIRK 555  
QY 519 YTEPRPIG 526  
Db 556 YI-PTNIG 562

## RESULT 15

US-09-022-949-2  
; Sequence 2, Application US/09022949  
; Patent No. 6187759  
; GENERAL INFORMATION:  
; APPLICANT: Tarpey, Ian  
; APPLICANT: Greenwood, Neil  
; TITLE OF INVENTION: Canine parvovirus DNA vaccination  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Akzo No. 6187759el Patent Dept.  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPC)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/022,949  
; FILING DATE: 11-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gormley, Mary E  
; REGISTRATION NUMBER: 34,409  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-948-7400  
; TELEFAX: 301-948-9751  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 584 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-022-949-2

Query Match 6.3%; Score 183; DB 3; Length 584;  
Best Local Similarity 20.4%; Pred. No. 3.7e-08;  
Matches 130; Conservative 91; Mismatches 223; Indels 192; Gaps 33;  
QY 5 GGAPMADNNEGADGVGNASGNHCDSTWLGDRVITST-----RTWALPTN-- 52  
Db 10 GGQP-AVRNERATGSGNGSGGGGGG--GGVGISTGTNNCTEPKFLENGWVEITANS 66  
QY 53 ----HL-----YKQI-----SSASTGASNDNHFGYSTPWGYDFNRFCHFSPRDW 95  
Db 67 RLVLNMPSENYRRVNVNLDKTAVNGNWLDDTHAQIVTPSLVDANAGWENFGDW 126  
QY 96 QRLNNMGRFKRLNFKLFNIOKVEV---TNDGVTTIANNLSTVQVFSSEYQLPYV 152  
Db 127 QLIVNTMSLHLVSEFEQEIFNVVLKTVSESAQPTKVYNNDLTASLMVALDSNNTMPT 186  
QY 153 LGSAGQGCL-----PPFA-----DVFMIPQGYLTLNNGSQVGRS----- 189

Db 187 PAARSETLGFYPMKPTIPTPMRYFQMDRTLIPSH-----TGTSGTPTNIYHGTDPD 240  
QY 190 -SPYCLE-YFPQMLRTGNP-TFSYTFEEVPHSHSAHSQSLDRMLNPLIDQLYLYLNR 246  
Db 241 VQFYTIENSVPVHLLRTGDEFATGTFFEDCKD--CRLTHTWQTNRALG--LPPFLNSLPQ 296  
QY 247 TQNS-----GSAQNK-----DLLFRSGSPAGMSVOPKNWLPGLPGPCYQQRV 287  
Db 297 SEGGTFGVIGVQODKRGVTOGMNTNITEATIMRPAEVGY-----APYYSFEAS 348  
QY 288 SK-----TKTDNNSNFTWQASKY-----NLNGRESIINPGT-----AMASHKD 327  
Db 349 TQGFPTPIAAGRGGAQTDENQAA---DGDPRYAFGRQHGQKTTTGTGETPERFTYIAHOD 405  
QY 328 DEDKFFPMGVMIFGKESAGASNTALD-NVMITDEEEIKATNPVATERFTGVAVAFQSSS 386  
Db 406 T-----GRYEGDWIQINFNLPVINDNVLPTDPIG-----GKTGINY--TN 446  
QY 387 TDPATGDVHAMGALPGVMWQDRDYLQGPWAKIPHTDGHFHPSPLMGGFGLK----- 439  
Db 447 IFNTYGLTLANNVP-----PVYPNGQIMDKFDTD-----LKPRLLHVNA 486  
QY 440 -----NPPQILIKNTVPVAPN---PPAEFSATKASFITQYSTQGVSVSEIEMELQKENS 490  
Db 487 PFVQCNCPGQLFVKVAPNLTNEDYDASANMSR-----IVTYSDFWWMKGKLVFKAKLRAS 542  
QY 491 KRWNPVEQVTSYAKSANVDFVDNNGLYTERPPIG 526  
Db 543 HTWNPIQMSIN-----IDNQFNVV-PSNIG 567

Search completed: January 21, 2004, 16:08:03  
Job time : 13.2857 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 16:04:27 ; Search time 24 Seconds  
(without alignments)  
4549.706 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGGAPMADNNEGADGVG.....NNGLYTEPRIGTRYLTRPL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*

2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*

3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*

4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*

5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*

6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*

7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*

8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*

9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*

10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*

11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*

12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*

13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*

14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*

15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*

16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*

17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2906	100.0	736	12	US-10-291-583-64
2	2906	100.0	736	12	US-10-423-704A-5
3	2884	99.2	736	12	US-10-291-583-65
4	2528	87.0	736	12	US-10-291-583-71
5	2528	87.0	736	12	US-10-423-704A-6
6	2481.5	85.4	533	14	US-10-038-972A-15
7	2481.5	85.4	598	14	US-10-038-972A-14
8	2481.5	85.4	735	12	US-10-291-583-70
9	2481.5	85.4	735	14	US-10-423-704A-4
10	2481.5	85.4	735	14	US-10-038-972A-13
11	2451.5	84.4	735	12	US-10-293-478-1
12	2451.5	84.4	735	12	US-10-291-583-68
13	2449.5	84.3	735	12	US-10-291-583-67
14	2448.5	84.3	735	12	US-10-291-583-66
15	2445.5	84.2	735	12	US-10-291-583-69

16	2441	84.0	728	12	US-10-291-583-106
17	2439	83.9	728	12	US-10-291-583-102
18	2437.5	83.9	731	12	US-10-291-583-89
19	2436	83.8	728	12	US-10-291-583-103
20	2436	83.8	728	12	US-10-291-583-107
21	2436	83.8	728	12	US-10-291-583-108
22	2434.5	83.8	738	12	US-10-291-583-93
23	2434.5	83.8	738	12	US-10-291-583-94
24	2431.5	83.7	731	12	US-10-291-583-88
25	2430	83.6	737	12	US-10-291-583-2
26	2430	83.6	737	12	US-10-423-704A-8
27	2429.5	83.6	738	12	US-10-291-583-79
28	2429.5	83.6	738	12	US-10-291-583-81
29	2428.5	83.6	738	12	US-10-291-583-85
30	2428.5	83.6	738	12	US-10-291-583-91
31	2427.5	83.5	738	12	US-10-291-583-92
32	2425	83.4	728	12	US-10-291-583-105
33	2423.5	83.4	733	12	US-10-291-583-87
34	2422.5	83.4	733	12	US-10-291-583-86
35	2422.5	83.4	738	12	US-10-291-583-80
36	2422	83.3	728	12	US-10-291-583-101
37	2421.5	83.3	733	12	US-10-291-583-90
38	2421.5	83.3	738	12	US-10-291-583-82
39	2419.5	83.3	729	12	US-10-291-583-110
40	2418.5	83.2	729	12	US-10-291-583-111
41	2418	83.2	728	12	US-10-291-583-104
42	2417	83.2	737	12	US-10-291-583-72
43	2415.5	83.1	738	12	US-10-291-583-84
44	2413	83.0	736	12	US-10-291-583-100
45	2413	83.0	736	12	US-10-423-704A-7

ALIGNMENTS

RESULT 1

US-10-291-583-64

Sequence 64, Application US/10291583

Publication No. US20030138772A1

GENERAL INFORMATION:

APPLICANT: Gao, Guangping

APPLICANT: Wilson, James M.

APPLICANT: Alvira, Mauricio

TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (

TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifi

FILE REFERENCE: UPN-02735USA

CURRENT APPLICATION NUMBER: US/10/291,583

CURRENT FILING DATE: 2002-11-12

PRIOR APPLICATION NUMBER: US 60/350,607

PRIOR FILING DATE: 2001-11-13

PRIOR APPLICATION NUMBER: US 60/341,117

PRIOR FILING DATE: 2001-12-17

PRIOR APPLICATION NUMBER: US 60/377,066

PRIOR FILING DATE: 2002-05-01

PRIOR APPLICATION NUMBER: US 60/386,675

PRIOR FILING DATE: 2002-06-05

NUMBER OF SEQ ID NOS: 120

SOFTWARE: PatentIn version 3.1

SEQ ID NO 64

LENGTH: 736

TYPE: PRT

US-10-291-583-64

Query Match 100.0%; Score 2906; DB 12; Length 736;

Best Local Similarity 100.0%; Pred. No. 2.2e-271;

Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGGAPMADNNEGADGVGNASGNHWCDSWLGDRVITTTSTRTWALPTNNHLYKQISS 60

Db 203 MASGGGAPMADNNEGADGVGNASGNHWCDSWLGDRVITTTSTRTWALPTNNHLYKQISS 262

QY 61 ASTGASNDNHYFGYSTPMGYFDNFRHCFHSPRDQRLNNNNWGRPKRLNPKLFNIQVK 120





Qy	181	NGSQAVERSSFYCLEYFPQSMLRGCNNFTFYTHTEEVPPHSSVAHSQSGLDRLANPLIDQY	240
Db	383	NGSQAVERSSFYCLEYFPQSMLRGCNNFTFYTHTEEVPPHSSVAHSQSGLDRLANPLIDQY	442
Qy	241	LYYLANTONQSSAONKOLLFSRSPAGMSVQPKNWLPGPCYQOORVSKTKTTONNSNFT	300
Db	443	LYYLANTONQSSAONKOLLFSRSPAGMSVQPKNWLPGPCYQOORVSKTKTTONNSNFT	502
Qy	301	WTGASKYNLNGRESIINPGTAMASHKODEDKFFPMSGVMIFGKESAGASALTALDNWIMTD	360
Db	503	WTGASKYNLNGRESIINPGTAMASHKDDKDKFFPMSGVMIFGKESAGASALTALDNWIMTD	562
Qy	361	ESEIKATNPVATEREGTVAVNFQSSSTDPAICGDVHANGALPGMWQORDVYLGPTIWAKI	420
Db	563	ESEIKATNPVATEREGTVAVNLQSSSTDPAICGDVHVWGALPGMWQORDVYLGPTIWAKI	622
Qy	421	PHTDGHFHPSPLMGGFGLKNPPPTLLIKNTVPVNPANPAEFSAIKFASFIQYSTGQSVBE	480
Db	623	PHTDGHFHPSPLMGGFGLKHPPPTLLIKNTVPVNPANPAEFSAIKFASFIQYSTGQSVBE	682
Qy	481	IWELOKENSKEWNDEVQVTSYNAKSAANDFTVDNNGLYTEPRPIGTRYLTRPL	534
Db	683	IWELOKENSKEWNDEVQVTSYNAKSAANDFTVDNNGLYTEPRPIGTRYLTRPL	736

RESULT 4  
 US-10-291-583-71  
 ; Sequence 71, Application US/10291583  
 ; Publication No. US20030138772A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gao, Guangping  
 ; APPLICANT: Wilson, James M.  
 ; APPLICANT: Alvira, Mauricio  
 ; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
 ; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifying  
 ; FILE REFERENCES: UPN-02735USA  
 ; CURRENT APPLICATION NUMBER: US/10/291.583  
 ; CURRENT FILING DATE: 2002-11-12  
 ; PRIOR APPLICATION NUMBER: US 60/350,607  
 ; PRIOR FILING DATE: 2001-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/341,117  
 ; PRIOR FILING DATE: 2001-12-17  
 ; PRIOR APPLICATION NUMBER: US 60/377,066  
 ; PRIOR FILING DATE: 2002-05-01  
 ; PRIOR APPLICATION NUMBER: US 60/386,675  
 ; PRIOR FILING DATE: 2002-06-05  
 ; NUMBER OF SEQ ID NOS: 120  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 71  
 ; LENGTH: 736  
 ; TYPE: PRT  
 ; ORGANISM: capsid protein of AAV serotype, clone AAV3  
 US-10-291-583-71

Query Match	87.0%;	Score	2528;	DB	12;	Length	736;
Best Local Similarity	85.8%;	Pred. No.	7.3e-235;				
Matches	459;	Conservative	28;	Mismatches	46;	Indels	2;
Gaps	2						
QY	1	MASGGGAPMADNNEGADGVGNASGNWHCDSTFWLGDRVITTTSTRTWALPTVNNHLYKQISS	60				
DB	203	MASGGGAPMADNNEGADGVGNSSGNWHCDSQWLGDRTVTTSTRTWALPTVNNHLYKQISS	262				
QY	61	ASTGASNDNHYFGYSTPWGYDFDNRFCHGSPRDWQRLINNNNGFRPRLNPKLFNIQVK	120				
DB	263	QS-GASNDNHYFGYSTPWGYDFDNRFCHGSPRDWQRLINNNNGFRPRLNPKLFNIQVR	321				
QY	121	EYTTDGVTTIANNLTSTVQVPSDSEYQLPVLGSAHQGCLPPFPADVFMIPQYGLTUN	180				
DB	322	GYTQNDGVTTIANNLTSTVQVFTTSEYQLPVLGSAHQGCLPPFPADVFMIPQYGLTUN	381				
QY	181	NGSQVGRSSFFCLEYFFPSQMLRTGNFTFSYTEEPFHSSYAHQSQDLRLANPLIDQY	240				

```

382  NGSQAVCRSFYCLYEYFPSPQMLRTGNNPQFSYTTEDVPFHSHSYAHQSQSLDRLMNPILDOY  441
241  LYLYLNRQ-QNSGSAQNKDILFSGSPAGMSVQPKNMLPGPCYRQORVSKTKTDDNNNSNF  299
442  LYLYLNRQTGTTGTTNQSRLLFSQAGPQMSLQARNMLPGPCYRQRLSKTANDNNNSNF  501
300  TWTGASKYNLNGRESIINPGTAMASHKDDKFFPMGSGWMTFGKESAGASNTALDNVMT  359
502  PWTAAASKYHLNGRDSLVPNGPAMASHKDDKFFPMHGNLIFGKEGTTASNAELDNVMT  561
360  DEBEIKATNPVATERFPGTVAVNFQSSSTDPAATGDVHAGALPGMWQDRDVLQGPWAK  419
562  DEBEIRTNPVATEQYGTGVANLQSSNAPTGTGVTHQAGALPGMWQDRDVLQGPWAK  621
420  IPTDGHFHPSPMLGGFGLKNPPQILIKNTFPVNPANPPAEFSATKFSFITYSTQGVSV  479
622  IPTDGHFHPSPMLGGFGLKHPPQIMIKNTFPVNPANPTTSPACFASFITQYSTQGVSV  681
480  EIEWELOKENSKRWNPEVQVTSNTAKSANVDFTVDNGLYTEPRDIGRILYLRPL  534
682  EIEWELOKENSKRWNPEIQVTSNTYKSNVNDFTVDITNGVSEPRDIGRILYLRNL  736

RESULT 5
US-10-423-704A-6
; Sequence 6, Application US/10423704A
; Publication No. US20030228282A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvir6, Mauricio
; TITLE OF INVENTION: Adeno-Associated Virus (AAV) Serotype 8 Sequences, Vectors
; TITLE OF INVENTION: Containing Same, and Uses Therefor
; FILE REFERENCE: UPN-02733AUSA
; CURRENT APPLICATION NUMBER: US/10/423,704A
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: US 60/341,151
; PRIOR FILING DATE: 2001-12-17
; CURRENT APPLICATION NUMBER: US 60/377,133
; PRIOR FILING DATE: 2002-05-01
; CURRENT APPLICATION NUMBER: US 60/386,122
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: PCT/US02/33630
; PRIOR FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 736
; TYPE: PRT
; ORGANISM: adeno-associated virus serotype 3
US-10-423-704A-6

```

Query Match	87.0%	Score	2528	DB 12	Length	736			
Best Local Similarity	85.8%	Pred.	No. 7.3e-235						
Matches	459	Conservative	28	Mismatches	46	Indels	2	Gaps	2

  

Qy	1	MASGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTTSTWALPTYNHLYKQISS	60
Db	203	MASGGGAPMADNNEGADGVGNSSGNHCHDSQWLGDRVITTTSTWALPTYNHLYKQISS	262
Qy	61	ASTGASNDNHYFGYSTPWGYFDNRFCHHFSPRDWORLIINNNGFRPRGLNPKLFNIOVK	120
Db	263	QS-GASNDNHYFGYSTPWGYFDNRFCHHFSPRDWORLIINNNGFRPKLSFKLFNIOVR	321
Qy	121	EVTITNDGVTITIANNLSTSTVQVFSDSYQYPYVLGSAHQGLCPPFPADVFMIPOYGYLTN	180
Db	322	GVTONDGTTIANNLSTSTVQVFTDSEYQLPYVLGSAHQGLCPPFPADVFVPOYGYLTN	381
Qy	181	NGSQAVGRSFCYCLEYFPSPQMLATGNPFTSYTFEEVPPHSSVAHSQSILDRANPLIDQY	240
Db	382	NGSQAVGRSFCYCLEYFPSPQMLATGNPFTSYTFEDVPPHSSVAHSQSILDRANPLIDQY	441
Qy	241	LYYLNRQTQ-NQSGAQNKDILLFRRSGPAGMSVQPKWLPGPCYRQQRVSXTKTDNNNSNF	299

Db 442 LYLNRQTGTTSGTTNQSKLLFSQAGPQMSLSQARNWLPQPCYRQRLSKTANDNNNSNF 501  
Qy 300 TWGASKYNLNGRESIINPGTAMASHKDDDEKFFPMGSMVIMFGKESAGASNTALDNVMT 359  
Db 502 PWTAAKTHLNGRDSLVPNGPAMASHKDDDEKFFPMHGNLIFGKEGTTASNAELDNVMT 561  
Qy 360 DEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAK 419  
Db 562 DEEIRITNPVATEQYGVVANNLQSNTPATTTGVNHQALPGMWQDRDVLQGPWAK 621  
Qy 420 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITYSTGQSV 479  
Db 622 IPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITYSTGQSV 681  
Qy 480 IEWELOKENSKRWNPEVQVTSNYAKSANVDFTVDNNGLYTEPRPIGRYLTRPL 534  
Db 682 IEWELOKENSKRWNPEIQYTSNYKSNVNDFTVDVTNGVISEPRPIGRYLTRNL 736

## RESULT 6

US-10-038-972A-15  
; Sequence 15, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038, 972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260,124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 533  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP3 capsid protien  
US-10-038-972A-15

Query Match 85.4%; Score 2481.5; DB 14; Length 533;  
Best Local Similarity 83.3%; Pred. No. 1.4e-230; Mismatches 37; Indels 1; Gaps 1;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
Qy 1 MASGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTTSTRTWALPTNNHLYKQISS 60  
Db 1 MATSGAPMADNNEGADGVGNSSGNHCHDSTWMDRVITTTSTRTWALPTNNHLYKQISS 60  
Qy 61 ASTGASNDNHFGYSTPWGYPDFNRFCHFSPRDQRLINNNWGRPKRLNPKLFNIQVK 120  
Db 61 QS-GASNDNHFGYSTPWGYPDFNRFCHFSPRDQRLINNNWGRPKRLNPKLFNIQVK 119  
Qy 121 EVTTNDGVTTIANLNTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTN 180  
Db 120 EVTQNDGTTIANLNTSTVQVTFDSEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTN 179  
Qy 181 NGSQAVGRSSFYCLEYFPSSQMLRTGNNTFTSYTPEVPFHSSYAHSSQSLDRLMNLIDQY 240  
Db 180 NGSQAVGRSSFYCLEYFPSSQMLRTGNNTFTSYTPEVPFHSSYAHSSQSLDRLMNLIDQY 239  
Qy 241 LYLNRQNSQSAQNKOLLFRRGSPAGMSVOPKMWLPQPCYRQORVSKTKTDNNNSNFT 300  
Db 240 LYLNRQNSQSAQNKOLLFRRGSPAGMSVOPKMWLPQPCYRQORVSKTKTDNNNSNFT 299  
Qy 301 WTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMGSMVIMFGKESAGASNTALDNVMT 360  
Db 300 WTGATKYNLNGRDSLVPNGPAMASHKDDDEKFFPMHGNLIFGKEGTTASNAELDNVMT 359  
Qy 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAK 420  
Db 360 EEEIRITNPVATEQYGVSVTNLQGRNQAAATADVNTQGVLPGMWQDRDVLQGPWAK 419  
Qy 421 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITYSTGQSV 480

Db 420 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITYSTGQSV 479  
Qy 481 IEWELOKENSKRWNPEVQVTSNYAKSANVDFTVDNNGLYTEPRPIGRYLTRPL 534  
Db 480 IEWELOKENSKRWNPEIQYTSNYKSNVNDFTVDVTNGVISEPRPIGRYLTRNL 533  
RESULT 7  
US-10-038-972A-14  
; Sequence 14, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038, 972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260,124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 598  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP2 capsid protien  
US-10-038-972A-14

Query Match 85.4%; Score 2481.5; DB 14; Length 598;  
Best Local Similarity 83.3%; Pred. No. 1.6e-230; Mismatches 51; Indels 1; Gaps 1;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
Qy 1 MASGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTTSTRTWALPTNNHLYKQISS 60  
Db 66 MATSGAPMADNNEGADGVGNSSGNHCHDSTWMDRVITTTSTRTWALPTNNHLYKQISS 125  
Qy 61 ASTGASNDNHFGYSTPWGYPDFNRFCHFSPRDQRLINNNWGRPKRLNPKLFNIQVK 120  
Db 126 QS-GASNDNHFGYSTPWGYPDFNRFCHFSPRDQRLINNNWGRPKRLNPKLFNIQVK 184  
Qy 121 EVTTNDGVTTIANLNTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTN 180  
Db 185 EVTQNDGTTIANLNTSTVQVTFDSEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTN 244  
Qy 181 NGSQAVGRSSFYCLEYFPSSQMLRTGNNTFTSYTPEVPFHSSYAHSSQSLDRLMNLIDQY 240  
Db 245 NGSQAVGRSSFYCLEYFPSSQMLRTGNNTFTSYTPEVPFHSSYAHSSQSLDRLMNLIDQY 304  
Qy 241 LYLNRQNSQSAQNKOLLFRRGSPAGMSVOPKMWLPQPCYRQORVSKTKTDNNNSNFT 300  
Db 305 LYLNRQNSQSAQNKOLLFRRGSPAGMSVOPKMWLPQPCYRQORVSKTKTDNNNSNFT 364  
Qy 301 WTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMGSMVIMFGKESAGASNTALDNVMT 360  
Db 365 WTGATKYNLNGRDSLVPNGPAMASHKDDDEKFFPMHGNLIFGKEGTTASNAELDNVMT 424  
Qy 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAK 420  
Db 425 EEEIRITNPVATEQYGVSVTNLQGRNQAAATADVNTQGVLPGMWQDRDVLQGPWAK 484  
Qy 421 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITYSTGQSV 480  
Db 485 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITYSTGQSV 544  
Qy 481 IEWELOKENSKRWNPEVQVTSNYAKSANVDFTVDNNGLYTEPRPIGRYLTRPL 534  
Db 545 IEWELOKENSKRWNPEIQYTSNYKSNVNDFTVDVTNGVISEPRPIGRYLTRNL 598

## RESULT 8

US-10-291-583-70  
; Sequence 70, Application US/10291583  
; Publication No. US20030138772A1

; GENERAL INFORMATION:  
 ; APPLICANT: Gao, Guangping  
 ; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
 ; FILE REFERENCE: US/10/291,583  
 ; CURRENT APPLICATION NUMBER: US 60/350,607  
 ; PRIOR FILING DATE: 2002-11-12  
 ; PRIOR APPLICATION NUMBER: US 60/341,117  
 ; PRIOR FILING DATE: 2001-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/377,066  
 ; PRIOR FILING DATE: 2002-05-01  
 ; PRIOR APPLICATION NUMBER: US 60/386,675  
 ; PRIOR FILING DATE: 2002-06-05  
 ; NUMBER OF SEQ ID NOS: 120  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 70  
 ; LENGTH: 735  
 ; TYPE: PRT  
 ; ORGANISM: capsid protein of AAV serotype, clone AAV2  
 US-10-291-583-70

Query Match 85.4%; Score 2481.5; DB 12; Length 735;  
 Best Local Similarity 83.3%; Pred. No. 2.3e-230;  
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
 QY 1 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVTTSTRTWALPTYNHLYKQISS 60  
 DB 203 MATSGAPMADNNEGADGVGNAGNWHCDSTWLGDRVTTSTRTWALPTYNHLYKQISS 262  
 QY 61 ASTGASNDNHFGYSTPMGYDFDNRFCHFSRDMQRLINNNWGRPKRLNFKLFNIOVK 120  
 DB 263 QS-GASNDNHFGYSTPMGYDFDNRFCHFSRDMQRLINNNWGRPKRLNFKLFNIOVK 321  
 QY 121 EVTTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQGLPPPPADVFMIPOYGYLTIN 180  
 DB 322 EVTQNDGTTIANNLTSTVQVDSSEYQLPYVLGSAHQGLPPPPADVFMIPOYGYLTIN 381  
 QY 181 NGSQAVGRSSFCYCLEYFPPSQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 240  
 DB 382 NGSQAVGRSSFCYCLEYFPPSQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 441  
 QY 241 LYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPQCYRQORVSKTKTDNNNSNFT 300  
 DB 442 LYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPQCYRQORVSKTKTDNNNSNFT 501  
 QY 301 WTGASKYNLNGRESINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNVMTD 360  
 DB 502 WTGASKYNLNGRESINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNVMTD 561  
 QY 361 EEEIKATNPVATERFGTVAVNFQSSSDPATGDVHAGMALPGMWQDRDVLQGPWAKI 420  
 DB 562 EEEIRTNVPATEQVSGSVTNLQGNRQAATADVNTQGVLPGMVWQDRDVLQGPWAKI 621  
 QY 421 PHTDGHFHPSPLMGGFGLKXPPQILIKNTVPANPPAEFSATKFAFSTQYSTQGSVE 480  
 DB 622 PHTDGHFHPSPLMGGFGLKXPPQILIKNTVPANPPAEFSATKFAFSTQYSTQGSVE 681  
 QY 481 IEWELQKSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
 DB 682 IEWELQKSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 735

RESULT 9  
 US-10-423-704A-4  
 ; Sequence 4, Application US/10423704A  
 ; Publication No. US20030228282A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gao, Guangping  
 ; APPLICANT: Wilson, James M.

; APPLICANT: Alvira, Mauricio  
 ; TITLE OF INVENTION: Adeno-Associated Virus (AAV) Serotype 8 Sequences, Vectors  
 ; FILE REFERENCE: US/10/423,704A  
 ; CURRENT APPLICATION NUMBER: US 60/341,151  
 ; PRIOR FILING DATE: 2001-12-17  
 ; PRIOR APPLICATION NUMBER: US 60/377,133  
 ; PRIOR FILING DATE: 2002-05-01  
 ; PRIOR APPLICATION NUMBER: US 60/386,122  
 ; PRIOR FILING DATE: 2002-06-05  
 ; PRIOR APPLICATION NUMBER: PCT/US02/33630  
 ; PRIOR FILING DATE: 2002-11-12  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 4  
 ; LENGTH: 735  
 ; TYPE: PRT  
 ; ORGANISM: adeno-associated virus serotype 2  
 US-10-423-704A-4

Query Match 85.4%; Score 2481.5; DB 12; Length 735;  
 Best Local Similarity 83.3%; Pred. No. 2.3e-230;  
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
 QY 1 MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVTTSTRTWALPTYNHLYKQISS 60  
 DB 203 MATSGAPMADNNEGADGVGNAGNWHCDSTWLGDRVTTSTRTWALPTYNHLYKQISS 262  
 QY 61 ASTGASNDNHFGYSTPMGYDFDNRFCHFSRDMQRLINNNWGRPKRLNFKLFNIOVK 120  
 DB 263 QS-GASNDNHFGYSTPMGYDFDNRFCHFSRDMQRLINNNWGRPKRLNFKLFNIOVK 321  
 QY 121 EVTTNDGVTIANNLTSTVQVDSSEYQLPYVLGSAHQGLPPPPADVFMIPOYGYLTIN 180  
 DB 322 EVTQNDGTTIANNLTSTVQVDSSEYQLPYVLGSAHQGLPPPPADVFMIPOYGYLTIN 381  
 QY 181 NGSQAVGRSSFCYCLEYFPPSQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 240  
 DB 382 NGSQAVGRSSFCYCLEYFPPSQMLRTGNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY 441  
 QY 241 LYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPQCYRQORVSKTKTDNNNSNFT 300  
 DB 442 LYLNRNTQSGSAQNKDLFSGSPAGMSVQPKNLPQCYRQORVSKTKTDNNNSNFT 501  
 QY 301 WTGASKYNLNGRESINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNVMTD 360  
 DB 502 WTGASKYNLNGRESINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNVMTD 561  
 QY 361 EEEIKATNPVATERFGTVAVNFQSSSDPATGDVHAGMALPGMWQDRDVLQGPWAKI 420  
 DB 562 EEEIRTNVPATEQVSGSVTNLQGNRQAATADVNTQGVLPGMVWQDRDVLQGPWAKI 621  
 QY 421 PHTDGHFHPSPLMGGFGLKXPPQILIKNTVPANPPAEFSATKFAFSTQYSTQGSVE 480  
 DB 622 PHTDGHFHPSPLMGGFGLKXPPQILIKNTVPANPPAEFSATKFAFSTQYSTQGSVE 681  
 QY 481 IEWELQKSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
 DB 682 IEWELQKSKRWNPVEQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 735

RESULT 10  
 US-10-038-972A-13  
 ; Sequence 13, Application US/10038972A  
 ; Publication No. US20020192823A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: J. Bartlett  
 ; TITLE OF INVENTION: AAV VECTORS AND METHODS  
 ; FILE REFERENCE: 28335/36996US  
 ; CURRENT APPLICATION NUMBER: US/10/038,972A  
 ; CURRENT FILING DATE: 2002-01-04

;; PRIOR APPLICATION NUMBER: US 60/260,124  
;; PRIOR FILING DATE: 2001-01-05  
;; NUMBER OF SEQ ID NOS: 19  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 13  
;; LENGTH: 735  
;; TYPE: PRT  
;; ORGANISM: adeno-associated virus 2 VPI capsid protien  
US-10-038-972A-13

Query Match 85.4%; Score 2481.5; DB 14; Length 735;  
Best Local Similarity 83.3%; Pred. No. 2.3e-230;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
QY 1 MASGGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTTSTRTWALPTVNNHLYKQISS 60  
DB 203 MATSGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTTSTRTWALPTVNNHLYKQISS 262  
QY 61 ASTGASNDNHFGYSTPWCYDFDNRFHCHFSRDMQRLINNNWGFPRKLNFKLFIQVK 120  
DB 263 QS-GASNDNHFGYSTPWCYDFDNRFHCHFSRDMQRLINNNWGFPRKLNFKLFIQVK 321  
QY 121 EVTTNDGVTIANNLTSTVQVFSSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 180  
DB 322 EVTQNDGTTIANNLTSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 381  
QY 181 NSQAVGRSSPYCLEYFPSPQMLRTGNFTFSYTFEEVPPHSSYAHQSOSLDRLMNPIDQY 240  
DB 382 NSQAVGRSSPYCLEYFPSPQMLRTGNFTFSYTFEEVPPHSSYAHQSOSLDRLMNPIDQY 441  
QY 241 LYLNRNTQNSGSAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQVRVSKTKTDNNNSFT 300  
DB 442 LYLNRNTQNSGSAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQVRVSKTKTDNNNSFT 501  
QY 301 WTGASKYNLNGRESIINPGTAMASHKODEDEKFPFMSGVMIIFGKESAGASNTALDNWITD 360  
DB 502 WTGATKYHLNGRDSLVPNGPAMASHKODEDEKFPFMSGVMIIFGKESAGASNTALDNWITD 561  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMVQDRDVLQGPWAKI 420  
DB 562 EEEIRTNVPATEQVGSVSTNLQRNQAAATADVNTQGVLPGMVQDRDVLQGPWAKI 621  
QY 421 PHTDGHFHPSPLMGFGKLNPPQILIKNTVPANPPAEFSAATKPFASITQYSTGQVSVE 480  
DB 622 PHTDGHFHPSPLMGFGKLNPPQILIKNTVPANPPAEFSAATKPFASITQYSTGQVSVE 681  
QY 481 IEWELQENSKRNWPEVOYTSYAKSANVDFTVDNNGLYTEPRPICTRYLTRPL 534  
DB 682 IEWELQENSKRNWPEVOYTSYAKSANVDFTVDNNGLYTEPRPICTRYLTRPL 735

RESULT 11  
US-10-293-478-1  
;; Sequence 1, Application US/10293478  
;; Publication No. US2003007841A1  
;; GENERAL INFORMATION:  
;; APPLICANT: PATEL, SALIL D.  
;; APPLICANT: MCARTHUR, JAMES G.  
;; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS  
;; FILE REFERENCE: 39672  
;; CURRENT APPLICATION NUMBER: US/10/293,478  
;; CURRENT FILING DATE: 2002-11-14  
;; PRIOR APPLICATION NUMBER: US/09/321,589  
;; PRIOR FILING DATE: 1999-05-28  
;; NUMBER OF SEQ ID NOS: 1  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1  
;; LENGTH: 735  
;; TYPE: PRT  
;; ORGANISM: Adeno-associated virus  
US-10-293-478-1  
Query Match 85.4%; Score 2481.5; DB 15; Length 735;

Best Local Similarity 83.3%; Pred. No. 2.3e-230;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
QY 1 MASGGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTTSTRTWALPTVNNHLYKQISS 60  
DB 203 MATSGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTTSTRTWALPTVNNHLYKQISS 262  
QY 61 ASTGASNDNHFGYSTPWCYDFDNRFHCHFSRDMQRLINNNWGFPRKLNFKLFIQVK 120  
DB 263 QS-GASNDNHFGYSTPWCYDFDNRFHCHFSRDMQRLINNNWGFPRKLNFKLFIQVK 321  
QY 121 EVTTNDGVTIANNLTSTVQVFSSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 180  
DB 322 EVTQNDGTTIANNLTSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVPMIPOYGYLTIN 381  
QY 181 NSQAVGRSSPYCLEYFPSPQMLRTGNFTFSYTFEEVPPHSSYAHQSOSLDRLMNPIDQY 240  
DB 382 NSQAVGRSSPYCLEYFPSPQMLRTGNFTFSYTFEEVPPHSSYAHQSOSLDRLMNPIDQY 441  
QY 241 LYLNRNTQNSGSAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQVRVSKTKTDNNNSFT 300  
DB 442 LYLNRNTQNSGSAQNKDLFLSRGSPAGMSVQPKNLPGPCYRQVRVSKTKTDNNNSFT 501  
QY 301 WTGASKYNLNGRESIINPGTAMASHKODEDEKFPFMSGVMIIFGKESAGASNTALDNWITD 360  
DB 502 WTGATKYHLNGRDSLVPNGPAMASHKODEDEKFPFMSGVMIIFGKESAGASNTALDNWITD 561  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMVQDRDVLQGPWAKI 420  
DB 562 EEEIRTNVPATEQVGSVSTNLQRNQAAATADVNTQGVLPGMVQDRDVLQGPWAKI 621  
QY 421 PHTDGHFHPSPLMGFGKLNPPQILIKNTVPANPPAEFSAATKPFASITQYSTGQVSVE 480  
DB 622 PHTDGHFHPSPLMGFGKLNPPQILIKNTVPANPPAEFSAATKPFASITQYSTGQVSVE 681  
QY 481 IEWELQENSKRNWPEVOYTSYAKSANVDFTVDNNGLYTEPRPICTRYLTRPL 534  
DB 682 IEWELQENSKRNWPEVOYTSYAKSANVDFTVDNNGLYTEPRPICTRYLTRPL 735

RESULT 12  
US-10-291-583-68  
;; Sequence 68, Application US/10291583  
;; Publication No. US20030138772A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Gao, Guangping  
;; APPLICANT: Wilson, James M.  
;; APPLICANT: Alvira, Mauricio  
;; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (A  
;; FILE REFERENCE: Sequences and Isolating No. US20030138772A1el Sequences Identifi  
;; CURRENT APPLICATION NUMBER: US/10/291,583  
;; CURRENT FILING DATE: 2002-11-12  
;; PRIOR APPLICATION NUMBER: US 60/350,607  
;; PRIOR FILING DATE: 2001-11-13  
;; PRIOR APPLICATION NUMBER: US 60/341,117  
;; PRIOR FILING DATE: 2001-12-17  
;; PRIOR APPLICATION NUMBER: US 60/377,066  
;; PRIOR FILING DATE: 2002-05-01  
;; PRIOR APPLICATION NUMBER: US 60/386,675  
;; PRIOR FILING DATE: 2002-06-05  
;; NUMBER OF SEQ ID NOS: 120  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 68  
;; LENGTH: 735  
;; TYPE: PRT  
;; ORGANISM: capsid protein of AAV serotype, clone A3.4  
US-10-291-583-68  
Query Match 84.4%; Score 2451.5; DB 12; Length 735;  
Best Local Similarity 82.6%; Pred. No. 1.8e-227;  
Matches 441; Conservative 37; Mismatches 55; Indels 1; Gaps 1;



Db 263 ES-GATNDNHFGYSTPWGDFNRFCHFSPRDQRLNNNNWGPFPKLNFKLFIQVK 321  
QY 121 EVTTNDGVTIANNLTSTVQVDSDEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTIN 180  
Db 322 EVTQNDGVTIANNLTSAVQVFTDSEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTIN 381  
QY 181 NGSQAVGRSSFYCLEYFPFQOMLRTGNFTFSYTFEVPFHSSYAHQSLSLDRMLNPLIDQY 240  
Db 382 NGSQAVGRSSFYCLEYFPFQOMLRTGNFTFSYTFEVPFHSSYAHQSLSLDRMLNPLIDQY 441  
QY 241 LYILNRTQNSGSAQNKDLFLFRGSPAGMSVQPKWMLPGCYRQORVSKTKTDNNNSNFT 300  
Db 442 LYILSKTOGTSQTQSRLOFQAGPSSMAQAQKWLPGPSYRQORMSKTANDNNSEFA 501  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMVMIFGKESAGASNTALDNVMTD 360  
Db 442 LYILSKTOGTSQTQSRLOFQAGPSSMAQAQKWLPGPSYRQORMSKTANDNNSEFA 501  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMVMIFGKESAGASNTALDNVMTD 360  
Db 502 WTAATKYYPNGRSLVNPFGPPMASHKDEDEKFFPMHGNLIFGKQGTGTTNVDIESVLITD 561  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGPFIWAKI 420  
Db 562 EEEIRTNPVATEQYQVATNRQONTTASVGSVDSQGLPGMWQDRDVLQGPFIWAKI 621  
QY 421 PHTDGHFHPSPMLMGFGFLKPNPPQIILIKNTVPANPPAFESATKFPASFTQYSTQVSVE 480  
Db 622 PHTDGHFHPSPMLMGFGFLKPNPPQIILIKNTVPANPPAFESATKFPASFTQYSTQVSVE 681  
QY 481 IEWELQKENSKRWNPEVQVTSYNAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 534  
Db 682 IEWELQKENSKRWNPEIQVTSNYSNVNVEFTVDANGVYSEPRPIGTRYLTRNL 735

RESULT 15  
US-10-291-583-69  
; Sequence 69, Application US/10291583  
; Publication No. US20030138772A1  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Guangping  
; APPLICANT: Wilson, James M.  
; APPLICANT: Alviria, Mauricio  
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus  
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifi  
; FILE REFERENCE: UPN-02735USA  
; CURRENT APPLICATION NUMBER: US/10/291,583  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR FILING DATE: 2001-12-17  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR FILING DATE: 2002-06-05  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 69  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: capsid protein of AAV serotype, clone A3.5  
US-10-291-583-69

Query Match 84.2%; Score 2445.5; DB 12; Length 735;  
Best Local Similarity 82.4%; Pred. No. 6.8e-227;  
Matches 440; Conservative 38; Mismatches 55; Indels 1; Gaps 1;  
QY 1 MASGGAPMADNNEGAGVGNASGNHCDSTWLGDRVITTTSTWALPTVNNHLYKQISS 60  
Db 203 MASGGAPMADNNEGAGVGNASGNHCDSTWLGDRVITTTSTWALPTVNNHLYKQISS 262  
QY 61 ASTGASNDNHFGYSTPWGDFNRFCHFSPRDQRLNNNNWGPFPKLNFKLFIQVK 120  
Db 263 ES-GATNDNHFGYSTPWGDFNRFCHFSPRDQRLNNNNWGPFPKLNFKLFIQVK 321  
QY 121 EVTTNDGVTIANNLTSTVQVDSDEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTIN 180

Db 322 EVTQNDGVTIANNLTSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVFMIPOYGYLTIN 381  
QY 181 NGSQAVGRSSFYCLEYFPFQOMLRTGNFTFSYTFEVPFHSSYAHQSLSLDRMLNPLIDQY 240  
Db 382 NGSQAVGRSSFYCLEYFPFQOMLRTGNFTFSYTFEVPFHSSYAHQSLSLDRMLNPLIDQY 441  
QY 241 LYILNRTQNSGSAQNKDLFLFRGSPAGMSVQPKWMLPGCYRQORVSKTKTDNNNSNFT 300  
Db 442 LYILSKTOGTSQTQSRLOFQAGPSSMAQAQKWLPGPSYRQORMSKTANDNNSEFA 501  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMVMIFGKESAGASNTALDNVMTD 360  
Db 502 WTAATKYYPNGRSLVNPFGPPMASHKDEDEKFFPMHGNLIFGKQGTGTTNVDIESVLITD 561  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGPFIWAKI 420  
Db 562 EEEIRTNPVATEQYQVATNRQONTTASVGSVDSQGLPGMWQDRDVLQGPFIWAKI 621  
QY 421 PHTDGHFHPSPMLMGFGFLKPNPPQIILIKNTVPANPPAFESATKFPASFTQYSTQVSVE 480  
Db 622 PHTDGHFHPSPMLMGFGFLKPNPPQIILIKNTVPANPPAFESATKFPASFTQYSTQVSVE 681  
QY 481 IEWELQKENSKRWNPEVQVTSYNAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 534  
Db 682 IEWELQKENSKRWNPEIQVTSNYSNVNVEFTVDANGVYSEPRPIGTRYLTRNL 735

Search completed: January 21, 2004, 16:18:25  
Job time : 25 secs